

**Repression of DNA-Binding-Dependent
Glucocorticoid Receptor-Mediated Gene Expression**

Thesis by
Katy Ann Muzikar

In Partial Fulfillment of the Requirements
for the Degree of
Doctor of Philosophy

California Institute of Technology
Pasadena, California

2011

(Defended November 29, 2010)

© 2011

Katy Ann Muzikar

All Rights Reserved



This thesis is dedicated to Jonathan...

...this journey has been ours together.

Acknowledgements

Five years seems like too short of a time to have learned all that I have learned, accomplished all that I have accomplished, and ultimately grown into the person and scientist that I find myself to be. I know that I wouldn't be nearly where I am today if it weren't for the love, support, encouragement, and mentoring of a great many people. I want to take time now to acknowledge some of those people, even knowing that I can't possibly properly thank everyone who has been a positive impact on my graduate career.

Peter: It has always been impressive to me that you are able to individually encourage each of your students along their own unique paths, and I will always be thankful for your positive energy, expertise, and patience as I learned to become the manager of my own “group of one.” I am confident in myself as a scientist and a teacher and am looking forward to going forth into this world to share with others the things that you have shared with me. *Jim:* Your energy, intelligence, and passion for the art of rational thought is inspiring. Your thoroughness and patience with me as a young Padawan has made my years in the Dervan Lab infinitely more productive than they would have been without your guidance. *Nick:* It has been an honor following in your footsteps. *Dan:* I have learned from you what it means to be a focused and dedicated scientist, and while I still have no real bark or bite, at least I learned from you how to growl, just a bit. *Michelle, Justin, and Mareike:* I love you like the little family we were—Mareike the wise and grounded older sister, Justin the hunky older brother that all my junior high school friends would have had a crush on, and Michelle, my crazy dizygotic twin (we aren't anything alike but we are linked together forever anyway). It was heartbreaking to watch each of you leave one by one, but the memories of the time we had together for coffee, Amigos, and fun in San Diego will last my whole lifetime. Thank you for getting me through the days when science refused to be kind, and celebrating with me when I finally got it right.

The Dervan Lab: In my five years I had the pleasure and good fortune of spending time with many great people: John, Anne, Ryan, Mike B., Jim S., Sherry, Christian, Carey, Dave C., Claire, Dan G., Ben, Dave M., Fei, Jordan, Jevgenij, JJ, Thomas, and Mandy. Every one of you has been a part of my Caltech experience—we have worked together to teach and learn from each other, to protect each other from the dangers of chemistry and of life, and to celebrate at the Rath Al Fresco whenever possible. The third floor of Church is home to an amazing group of people—ever changing yet always a group with a cooperative spirit and passion for life. Best wishes to each of you in your own individual journeys and I can't wait to see each of you whenever life throws us back together again.

Thank you to my thesis committee: Linda Hsieh-Wilson, Doug Rees, and Dennis Dougherty. Your advice and support has been invaluable and it has been an honor to have scientific and career discussions with such brilliant minds.

There is absolutely no way I would have accomplished everything in this thesis without the services of the friendly, dedicated, amazing support staff of the Division of Chemistry and Chemical Engineering. Joe Drew and Ron Koen, I hope that you guys know that you made my day infinitely better countless times—with banter, with problem-solving, and of course with sugar. Paul Carroad, the go-to guy for absolutely everything, I'm not sure that anything would have gotten done without your help. Lynne Martinez, you have been an angel of patience and a bright smile of encouragement when I needed it most. Dianne Buchness, Anne Penny, and Agnes Tong, who took care of so many behind-the-scenes things that I don't even know about to make sure that my time as a graduate student could be productive and (somewhat) less full of stress. Steve Gould, Chris Smith and Leah Mentch, who put up with last-minute-radiation-orders and everything else that I brought them and always did so without complaint. Much thanks to the group that fixes everything we break, no matter how many times we break it: Tom Dunn, Rick Gerhart, Tony Solyom,

Mike Roy, and Steve Olson. The excellent facilities and great intellect of Mona Shahgholi, David Vander Velde, Scott Ross and Scott Virgil have been invaluable.

Not only is Caltech top-class in research, but our support staff and departments are excellent. To the Department of Environmental Health and Safety, especially Haick Issaian, Karen Baumgartner, Larry Martinez, Art Seiden, and Andrea Acosta thank you for your continued commitment to keep us safe, to keep the community safe, and to continue to educate us to be better and safer scientists. Karen--it has been wonderful having radiation delivered by such a vibrant personality, and best wishes in life to you and your family. A special thank you to the HVAC technicians, especially Steve and Reuben, for bandaging our McQuay units time and time again (even at the cost of getting soaking wet).

A great big amazing thank you to my family. My mother and father are responsible for my deep-rooted love of science. They took me to the Exploratorium and I never looked back. Mom--your love and support, both emotional and physical, has meant the world to me, and I don't know that I would have made it through this time without our near-daily phone conversations. Dad--your wonder at the natural world around you must be in your genes, because you definitely passed it down to me. I'm glad that I could be the organic chemist that you always wanted to be. My grandparents have always been a source of inspiration and having their support and interest in my work is an amazing feeling. To the Josephson clan, yeah, you know who you are, your energy and enthusiastic support of anything and everything I do is astounding. I couldn't have picked a better group of people to add to my family. To Tricksy, your unconditional love and warm cuddly support has made my world a better place.

And finally to Jonathan: Thank you for being strong enough, and for sharing your strength. I love you.

Abstract

Gene expression is controlled by transcription factors that regulate the rates at which genes are expressed either by recruiting or inhibiting protein complexes that bind to the promoters or enhancers of target genes. Molecules that can specifically modulate these protein-DNA interfaces show promise as tools for understanding gene regulation pathways and may have application in human medicine. Hairpin pyrrole-imidazole polyamides are programmable oligomers that bind the DNA minor groove in a sequence-specific manner with affinities comparable to those of natural DNA-binding proteins. These cell-permeable small molecules have been shown to enter the nuclei of live cells, disrupt protein-DNA interactions, and downregulate endogenous gene expression. This thesis describes the use of polyamides to modulate gene expression in order to probe gene regulation mechanisms of several different biologically relevant systems. A polyamide is designed to target the glucocorticoid receptor transcription factor DNA binding site located in the promoter of the glucocorticoid-induced leucine zipper gene. This polyamide is shown to bind with high affinity to the promoter sequence, modulate the expression of this gene, and disrupt the binding of the protein to the gene's promoter. Examination of the global effects of this polyamide on mRNA transcription is used to elucidate a list of genes that are regulated by a glucocorticoid receptor protein-DNA dependent mechanism. Also in this thesis, the specificities of a Cy3-labeled polyamide known to downregulate expression of the Vascular Endothelial Growth Factor is examined using DNA microarrays composed of hairpins harboring all 524,800 unique 10 base pair DNA sequences. We experimentally verify the correlation of Cy3 fluorescence intensity with quantitative DNase I footprint-derived binding affinities. Additionally, progress is made towards the polyamide-mediated inhibition of Myc/Max transcription factor gene regulation.

Table of Contents

Acknowledgements.	iv
Abstract.	vii
Table of Contents	viii
List of Figures and Schemes.	x
List of Tables.	xiii
List of Abbreviations.	xiv
List of Symbols and Nomenclature	xviii
Chapter 1: Introduction.	20
1.1 Gene expression and the central dogma	21
1.2 Regulation of gene expression: transcriptional modulation	22
1.3 Molecular recognition of DNA	24
1.4 Sequence-specific DNA-binding small molecules	26
1.5 Polyamide inhibition of gene regulation in live cell culture.	29
1.6 An allosteric model for polyamide inhibition of steroid hormone receptors	32
1.7 Scope of this work	37
1.8 References.	38
Chapter 2: Repression of DNA-Binding-Dependent Glucocorticoid Receptor-Mediated Gene Expression	41
Abstract.	42
2.1 Introduction.	43
2.2 Results.	48
2.3 Discussion.	55
2.4 Materials and Methods	60
2.5 Acknowledgements.	68
2.6 References.	73

Chapter 3: Quantitative Microarray Profiling of DNA-Binding Molecules	76
Abstract	77
3.1 Introduction	78
3.2 Results	81
3.3 Discussion	97
3.4 Materials and Methods	101
3.5 Acknowledgements	107
3.6 References	108
Chapter 4: Progress Towards Polyamide Inhibition of Myc-Activated Gene Expression by Antagonism of the E-box Fragment 5'-WCGGW-3'	112
Abstract	113
4.1 Introduction	114
4.2 Experimental Design	118
4.3 Closing Remarks	134
4.4 Materials and Methods	135
4.5 References	149
Appendix A: Progress Towards Incorporation of Furan Rings into Pyrrole-Imidazole Polyamides	154
A.1 Introduction	155
A.2 Experimental Design	158
A.3 Future Directions	163
A.4 Synthetic Efforts	163
A.5 Materials and Methods	165
A.6 References	176

List of Figures and Schemes

Chapter 1

Figure 1.1 X-ray crystal structures of transcription factor–DNA complexes	23
Figure 1.2 The structure of DNA	24
Figure 1.3 Major and minor groove hydrogen bonding patterns of the four Watson-Crick base pairs	25
Figure 1.4 Chemical structures of DNA-binding small molecule natural products . .	26
Figure 1.5 X-ray crystal structures of distamycin bound to DNA	27
Figure 1.6 Recognition of the DNA minor groove by polyamides.	29
Figure 1.7 Polyamides as regulators of gene expression in cell culture	30
Figure 1.8 Atomic model of the cooperative assembly of interferon- β enhancesome	33
Figure 1.9 Comparison of native DNA to polyamide/DNA complex.	35

Chapter 2

Figure 2.1 Crystallographic structures of the glucocorticoid receptor	43
Figure 2.2 Simplified model for the response of glucocorticoid receptor to steroid hormone stimulus	44
Figure 2.3 Effect of polyamide-DNA binding on GR gene regulation.	46
Figure 2.4 Polyamide design and <i>GILZ</i> promoter structure	47
Figure 2.5 DNase I footprinting of <i>GILZ</i> promoter region	49
Figure 2.6 <i>In vitro</i> displacement of GR binding by polyamide 1	50
Figure 2.7 Inhibition of dexamethasone-induced <i>GILZ</i> expression by 1 and 2	52
Figure 2.8 Global effects of polyamides on GR-regulated genes.	53
Figure 2.9 Timeline of cell treatment protocol, used for RT-PCR, ChIP and microarray assays	65

List of Figures and Schemes

Chapter 3

Figure 3.1 Methods for analyzing DNA binding specificity.	80
Figure 3.2 Polyamides for CSI Studies.	82
Figure 3.3 Histogram of microarray intensities	83
Figure 3.4 Histogram of microarray fractional standard deviations.	84
Figure 3.5 Insert sequences utilized in plasmids	85
Figure 3.6 DNase I footprinting gels and corresponding isotherms of polyamides 1 and 2 on pKAM3 and pKAM4.	87
Figure 3.7 DNase I footprinting gels and corresponding isotherms of polyamides 3 and 4 on pJWP17	88
Figure 3.8 CSI array intensities correlate well with DNase I footprinting- determined K_a values	90
Figure 3.9 Correlation of footprinting and CSI data.	91
Figure 3.10 Cy3-labeled polyamides and unlabeled polyamides correlate well	92
Figure 3.11 K_a -weighting components of individual sequence logos does not alter the sequence logo	93

Chapter 4

Figure 4.1 The Myc-Max transcription factor	114
Figure 4.2 Polyamide inhibition of Myc-Max DNA binding	116
Figure 4.3 Initial polyamide library.	118
Figure 4.4 Library of compounds synthesized by K. Muzikar and D. Harki for biological studies	120
Figure 4.5 Binding studies of polyamides 7 and 8	122
Figure 4.6 Fluorescein isothiocyanate (FITC) conjugate polyamides synthesized by K. Muzikar and D. Harki	123
Figure 4.7 Cellular uptake of polyamides 13 and 14	124

List of Figures and Schemes

Figure 4.8 Cellular uptake of polyamides 13-16	124
Figure 4.9 Schematic representation of E-box locations relative to transcription start site of eIF4E.	127
Figure 4.10 Doxycycline-induced Myc and eIF4E expression in MCF7-cl35 cells	130
Figure 4.11 siRNA against Myc in NCI-H82 cells	131
Figure 4.12 qRT-PCR results in NCI-H82 cells	133
Scheme 4.1 Synthesis of polyamide 11	137
Scheme 4.2 Synthesis of cyclic polyamide 12	139
Appendix A	
Figure A.1 DNA-binding molecules with furan heterocycle constituents	156
Figure A.2 Furan monomers incorporated into 5'-WGGWCW-3'targeted polyamide scaffold	158
Figure A.3 Py-Im-Fn polyamide library designed to test specificity and lipophilicity	159
Scheme A.1 Attempted synthesis by D. Gubler of furan monomer Fn.	164
Scheme A.2 Synthesis of oxazole monomer Ox by D. Gubler.	165
Scheme A.3 Synthesis of Boc-protected Fn monomer as reported by Süssmuth. . .	166
Figure A.4 ¹ H NMR (500 MHz, DMSO) of compound 14	170
Figure A.5 ¹³ C NMR (500 MHz, DMSO) of compound 14	171

List of Tables

Chapter 2

Table 2.1 Genes affected >2-fold by dexamethasone and mifepristone whose activity is modulated by polyamide 1 and not by polyamide 2	54
Table 2.2 Genes identified in our study that have previously been shown to have GR-occupied and/or functional GREs.	59
Table 2.3 Microarray fold changes under all treatment conditions of sequences affected ≥ 2 -fold by dexamethasone and mifepristone.	69

Chapter 3

Table 3.1 Quantitative DNase I footprinting-derived K_a values (M^{-1}) for 1 and 2 . .	86
Table 3.2 Quantitative DNase I footprinting-derived K_a values (M^{-1}) for 1 and 2 . .	88
Table 3.3 Microarray-derived binding affinities and specificities of all single base pair mismatch sites for polyamide 2	95
Table 3.4 Microarray-derived binding affinities and specificities of all single base pair mismatch sites for polyamide 4	96

Chapter 4

Table 4.1 Assessment of binding affinities and specificities of polyamides 1-5 . . .	119
Table 4.2 Cellular localization of polyamide–dye conjugates in cultured cells . . .	125
Table 4.3 E-box and surrounding DNA sequences of several known direct Myc target genes.	126
Table 4.4 Thermal melting temperature studies of polyamides 7 and 11 on eIF4e E-box 1	128

Appendix A

Table A.1 Assessment of binding affinities and specificities of compounds 1, 4, 5, and 7	160
Table A.2 LogD at pH 7.4 of compounds 1, 4, 5, and 7 at 0.1 mM concentration . .	162

List of Abbreviations

[PA]	free polyamide concentration
A	adenine
Å	angstrom
Ac	acetyl
Ac ₂ O	acetic anhydride
ADMET	absorption, distribution, metabolism, excretion and toxicity
A ₅₉₅	absorbance maximum
AP-1	activator protein 1
AR	androgen receptor
ARE	androgen response element
A·T	adenine Watson-Crick hydrogen bonded to thymine
ATCC	American Type Culture Collection
atm	atmosphere
β	beta-amino alanine
Boc	<i>tert</i> -butoxycarbonyl
Boc-Im-OH	(4-[(<i>tert</i> -Butoxycarbonyl)amino]-1-methylimidazole-2-carboxylic acid)
Boc ₂ O	di- <i>tert</i> -butyl dicarbonate
Boc-Py-OBt	[(1,2,3-Benzotriazol-1-yl 4-[(<i>tert</i> -Butoxycarbonyl)amino]-1-methylpyrrole-2-carboxylate)
bp	base pair
BSA	bovine serum albumin
°C	degrees Celsius
C	cytosine
cDNA	complementary deoxyribonucleic acid
C·G	cytosine Watson-Crick hydrogen bonded to guanine
calc'd	calculated
Cbz	carbobenzyloxy
ChIP	chromatin immunoprecipitation
cm	centimeter
CSI	Cognate Site Identity/Cognate Site Identifier
Ct	2-carboxy-3-chlorothiophene
CT-FBS	charcoal treated fetal bovine serum
Cy3	Cyanine 3
Da	Dalton
dATP	2'-deoxyadenosine triphosphate
DABA	diaminobutyric acid
DCM	dichloromethane
dex	dexamethasone
DFO	deferoxamine

continued on page xv

List of Abbreviations

DHT	dihydrotestosterone
DIEA	N,N-diisopropylethylamine
DMF	N,N-dimethylformamide
DMSO	dimethylsulfoxide
DNA	deoxyribonucleic acid
dox	doxycycline
Dp	N,N-dimethylaminopropylamine
DPPA	diphenylphosphoryl azide
DTT	dithiothreitol
EDTA	ethylenediaminetetraacetic acid
eIF4e	eukaryotic translation initiation factor 4E
ELISA	enzyme-linked immunosorbant assay
EMSA	electrophoretic mobility shift assay
ESI	electrospray ionization
Et ₂ O	diethyl ether
Ex	excitation
FBS	fetal bovine serum
FITC	fluorescein isothiocyanate
Fmoc	fluorenylmethyloxycarbonyl
Fn	4-aminofuran
γ-DABA	γ-2,4-diaminobutyric acid
G	guanine
G·C	guanine Watson-Crick hydrogen bonded to cytosine
GABA	γ-aminobutyric acid
GC	glucocorticoid
GILZ	glucocorticoid-induced leucine zipper
GR	glucocorticoid receptor
h	hour(s)
HIF-1α	hypoxia inducible factor 1α
Hp	3-hydroxypyrrole
HPLC	high-performance liquid chromatography
HRE	hypoxic response element
HRP	horseradish peroxidase
Hsp	heat-shock protein
IgG	immunoglobulin G
Im	N-methylimidazole
IPA	isophthalic acid
IPTG	isopropyl β-D-1-thiogalactopyranoside
k	kilo (1 x 10 ³)

continued on page xvi

List of Abbreviations

K_a	association constant
K_d	dissociation constant
λ	wavelength
LN_2	liquid nitrogen
m/z	mass to charge ratio
μ	micro (1×10^{-6})
M	molar
m	milli (1×10^{-3})
Max	Myc associated protein X
max	maximum
MALDI	matrix-assisted LASER desorption/ionization
MAS	maskless array synthesis
mif	mifepristone
min	minute(s)
mol	mole(s)
MPE	methidiumpropyl-ethylenediaminetetraacetic acid
mRNA	messenger ribonucleic acid
MS	mass spectrometry
N	A, T, G, or C
n	nano (1×10^{-9})
n-BuLi	n-butyl lithium
NF- κ B	nuclear factor- κ B
Θ	fractional occupancy
OBt	hydroxytriazole ester
p	pico (1×10^{-12})
PCR	polymerase chain reaction
PET	Paired End diTag
PIC	protease inhibitor cocktail
PMSF	phenylmethanesulfonylfluoride
PSA	prostate-specific antigen
Py-Im	pyrrole-imidazole
qPCR	quantitative polymerase chain reaction
RT	room temperature
RT-PCR	reverse transcriptase polymerase chain reaction
PAGE	polyacrylamide gel electrophoresis
PBS	phosphate-buffered saline
Py	N-methylpyrrole
PyBOP	(benzotriazol-1-yloxy)tripyrrolidinophosphonium hexafluorophosphate

continued on page xvii

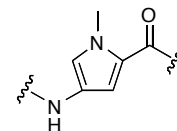
List of Abbreviations

R	guanine or adenine
RCF	relative centrifugal force
RIPA	radio immunoprecipitation assay
RNA	ribonucleic acid
RNAi	ribonucleic acid interference
RT	reverse transcription
SAGE	serial analysis of gene expression
s-BuLi	<i>sec</i> -butyl lithium
siRNA	small interfering ribonucleic acid
Smad	Sma and Mad-related protein
STAT	signal transduction and activator of transcription
T	thymine
T·A	thymine Watson-Crick hydrogen bonded to adenine
t-BuOH	<i>tert</i> -butanol
TF	transcription factor
TFA	trifluoroacetic acid
TFO	triplex-forming oligonucleotides
THF	tetrahydrofuran
T _m	midpoint of transition temperature
TOF	time-of-flight
TFRE	transcription factor response element
tri/triamine	3,3'-diamino-N-methyldipropylamine
U	uracil
UV	ultraviolet
VEGF	vascular endothelial growth factor
Vis	visible
W	adenine or thymine
X-gal	bromo-chloro-indolyl-galactopyranoside

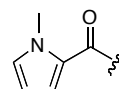
List of Symbols and Nomenclature



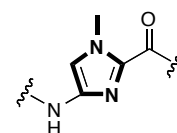
-Py-



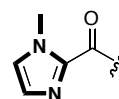
Py-



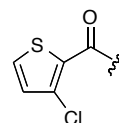
-Im-



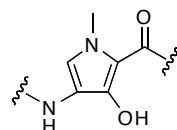
Im-



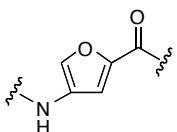
Ct-



-Hp-



-Fn-

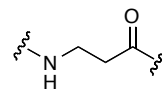


continued on page xix

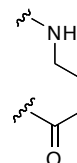
List of Symbols and Nomenclature



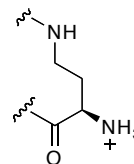
-β-



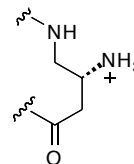
-γ-



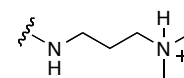
-γ-DABA-



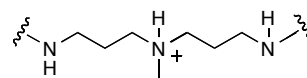
-(R)-β-amino-γ-



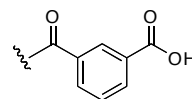
-Dp



-tri-



-IPA



Chapter 1

Introduction

1.1 Gene expression and the central dogma

In 1958 the means by which life as we know it persists and thrives was enumerated in the form of the central dogma of molecular biology.¹ In its simplest form, the dogma states that each living cell holds the blueprint to life in the form of genes, and that this genomic information is transcribed from DNA to RNA, which is then translated into proteins. Over half a century later, innumerable advances in technology have expanded this simple view to include a host of mechanisms by which each step of this process is influenced and modulated in response to cellular needs. The genome is not a blueprint of a static cellular state but actually encompasses all the instructions needed for a cell to respond to environmental stimuli. Sequencing of the human genome estimates that humans possess 20,000 to 30,000 genes,^{2,3} all of which are maintained folded within chromosome structures and are accessed as needed by the cell to maintain normal function. Information encoding the structure, regulation, and expression of each of these genes as well as their expressed products is encoded in the nucleotide sequences of the base pairs of the DNA molecules that make up the genome.

Gene expression is the process by which a gene's coded information is converted from these nucleotide sequences into functioning gene products. This process is used by all known life forms to generate the RNAs, proteins, and various macromolecular machineries needed for life. In eukaryotic cells, gene expression is a complex process involving a variety of steps prior to the actual synthesis of a protein. These include the transcription of the gene into the primary RNA product, processing of this initial gene transcript to remove intron sequences and create the mature 3' terminus, transport of the processed mRNA transcript to the cytoplasm, and then, finally, translation of the messenger RNA into protein. With very few exceptions, all of the genes that encode proteins follow this pathway.⁴

1.2 Regulation of gene expression: transcriptional modulation

Any step of the gene expression process is subject to modulation, from the transcription of DNA into RNA all the way through the post-translational modification of the final protein. Every cell in any given organism contains the exact same genes, yet the multitude of cell types in a human body each display a different phenotype, produce different proteins at different stages of the life cycle, and somehow display different programs of gene expression. At the most fundamental level, regulation of gene expression is what allows genotype to be converted into phenotype. Modulation of the amount and timing of the appearance of the final gene product allows the cell to maintain control over structure and function and provides the basis for the versatility and adaptability of a living organism. A cell's ability to regulate and control the expression of each gene product allows the cell the flexibility to respond to environmental stimuli such as chemical signaling, environmental variation, cell damage, etc.

Regulation at the transcriptional level is a key element in the regulation of gene expression and can be divided into three general categories of influence: 1) Regulation at the genetic level involves direct interaction of a control factor with the gene in question, 2) Regulation via interaction of a control factor with the transcription machinery that results in modulation of gene expression, and 3) Epigenetic regulation involving global alterations in DNA structure that influence transcription. Direct interaction with DNA is the simplest and most direct method by which a protein can modulate the level of gene transcription. In general, genes harbor several protein binding sites specifically utilized for the regulation of transcription. These protein binding sites include enhancers, insulators, repressors, and silencers and the mechanisms for regulating transcription through these binding sites are complex and vary from activating transcription by recruiting RNA polymerase to blocking key RNA polymerase binding sites to hinder gene transcription.

Transcription Factors

Transcription factor proteins constitute a major class of DNA-binding molecules that participate in the transcriptional control of gene expression. Transcription factors bind to DNA in gene promoters or enhancer regions, generally upstream of the transcription start site, and modulate the frequency of RNA polymerase binding to the gene and subsequent transcription. Transcription factors vary widely in architecture, as the small sample of transcription factors shown in **Figure 1.1** illustrates. While different transcription factors accomplish their tasks via different means, all transcription factors display two critical abilities—the ability to bind to DNA in a sequence-specific manner, and the ability to cause a modulation of transcription once bound.⁴ Most transcription factors possess a specific DNA-binding domain that is responsible for recognition and interaction with a specific DNA sequence, typically four to eight nucleotides in length, often located in an enhancer or promoter region of a given gene.

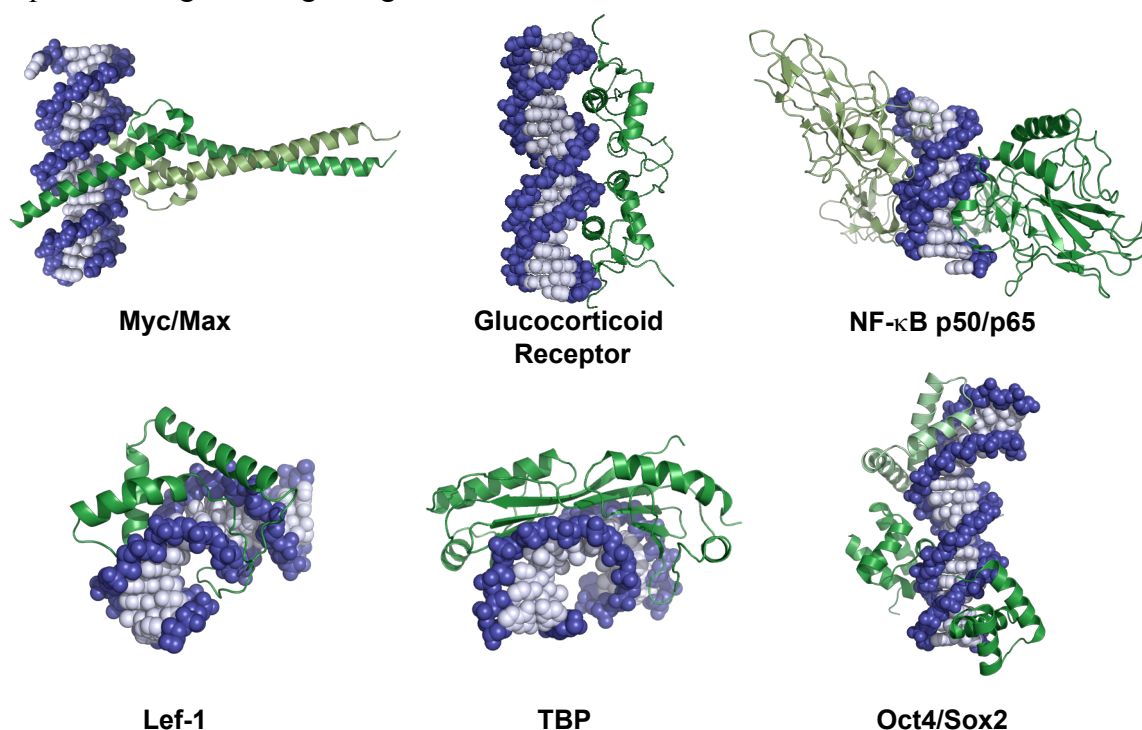


Figure 1.1 X-ray crystal structures of transcription factor–DNA complexes

Myc/Max (PDB 1NKP),⁵ *Glucocorticoid Receptor* (PDB 1R40),⁶ *NF-κB* (PDB 1LE5),⁷ *Lef-1* (PDB 2LEF),⁸ *TBP* (PDB 1TGH),⁹ *Oct4/Sox2* (PDB 1O4X).¹⁰

1.3 Molecular recognition of DNA

Deoxyribonucleic acid (DNA):

A single strand of DNA consists of four different bases linked by a phosphodiester deoxyribose sugar backbone. Two strands of DNA intertwine to form a double helix associated by hydrogen bonds between the Watson-Crick base pairs such that thymine (T) pairs with adenine (A), and cytosine (C) pairs with guanine (G) (**Figure 1.2**). In a DNA double helix the direction of the nucleotides in one strand is opposite to their direction in the other strand, thus the strands are antiparallel. The helix of B-form DNA, the average conformation adopted by the majority of biologically active DNA sequences, is right handed

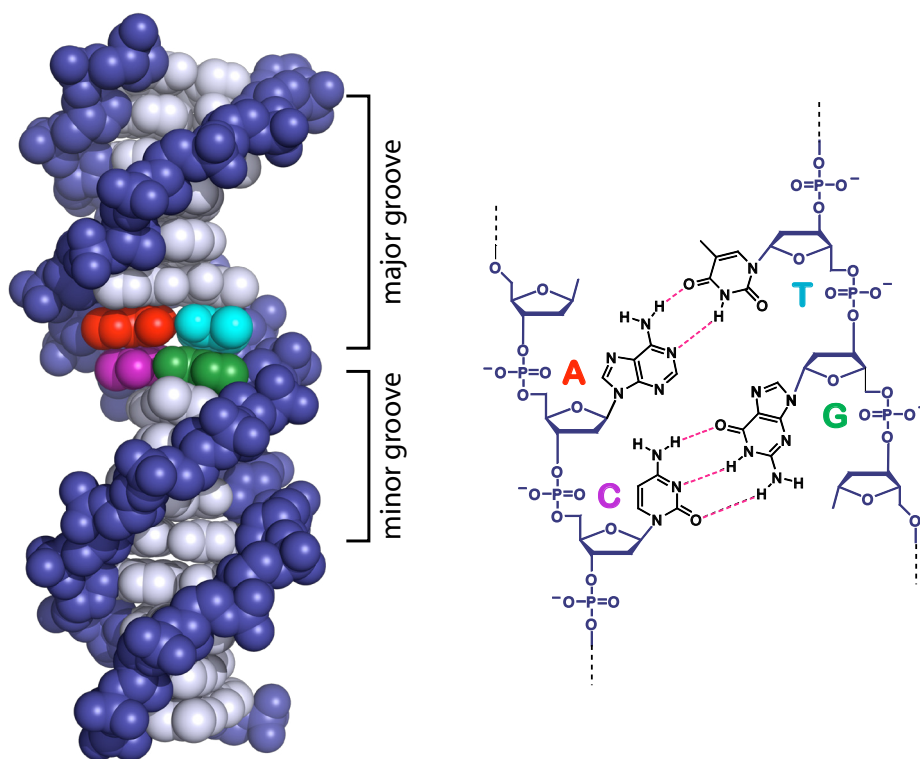


Figure 1.2 The structure of DNA

Left) X-ray crystal structure of B-form DNA. The phosphodiester-linked deoxyribose backbone is shown in blue, and the Watson-Crick base pairs are shown in gray (PDB 1BNA).¹¹ Highlighted: adenine (red) thymine (blue) cytosine (magenta) and guanine (green). Right) Chemical structures of phosphodiester-linked hydrogen-bonded base pairs. Adenine (A) is bonded to thymine (T) and cytosine (C) is bonded to guanine (G) Dashed lines indicate hydrogen bonds.

and displays ten base pairs per turn with the plane of each hydrogen-bonded base pair lying perpendicular to the helical axis. This helix is not symmetrical, instead a wide major groove and a narrow minor groove line the helix and DNA sequences can be distinguished by the pattern of functional groups, e.g., hydrogen bond donors and acceptors, displayed on the edges of the base pairs in either groove.

Sequence-specific recognition of DNA:

Proteins and small molecules can recognize B-form DNA through interactions with the major groove, minor groove, and phosphate backbone, or a combination of these elements. These interactions can be mediated through electrostatics, hydrogen bonding, van der Waals interactions, and, in the case of intercalators, via base pair stacking. The DNA base pair edges in the major groove and minor groove provide a sequence-specific array of functionality for hydrogen bonding, hydrophobic interactions, and steric complementarity with proteins and small molecule binders.¹²⁻¹⁶ It is these patterns of hydrogen donors and acceptors that allow for sequence-specific recognition of DNA by proteins and small

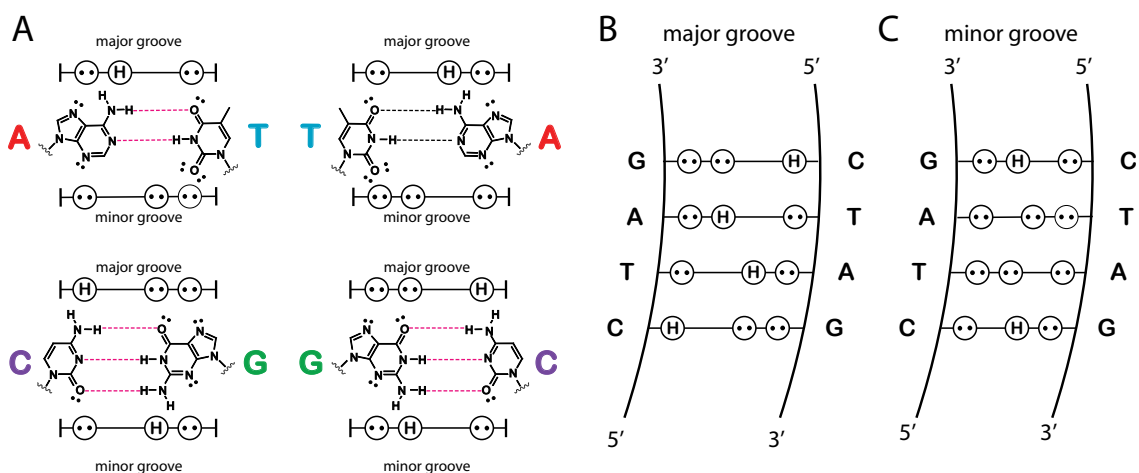


Figure 1.3 Major and minor groove hydrogen bonding patterns of the four Watson-Crick base pairs

Circles with dots represent lone pairs on purine N3 or pyrimidine O2, and circles with an H represent the 2-aminogroup of guanine (G-NH2). A) Major and minor groove hydrogen bonding patterns of each base pair B) Hydrogen bonding pattern offered to the major groove by the sequence 5'-GTAC-3'. C) Hydrogen bonding pattern offered to the minor groove by the sequence 5'-GTAC-3'

molecules that can form hydrogen bonds with the functional groups present in the grooves. In a formal sense, the four Watson-Crick base pairs can be differentiated on the groove floor by the specific positions of hydrogen bond donors and acceptors, as well as by subtle differences in molecular shape. **Figure 1.3** depicts the differences in these hydrogen bond donors and acceptors as seen on the floor of each groove.

1.4 Sequence-specific DNA-binding small molecules

In addition to DNA-binding proteins such as the transcription factors described above, there are also small molecule natural products that recognize and bind specific DNA sequences. Four particular examples are shown in **Figure 1.4**: calicheamicin oligosaccharide, chromomycin, actinomycin D, and distamycin A. Calicheamicin oligosaccharide has been shown to recognize and bind the minor groove monomerically at 5'-TCCT-3' sequences.¹⁷ Chromomycin targets the sequence 5'-GGCC-3' and binds in the minor groove of DNA in a 2:1 ligand:DNA stoichiometry. Chromomycin's biological

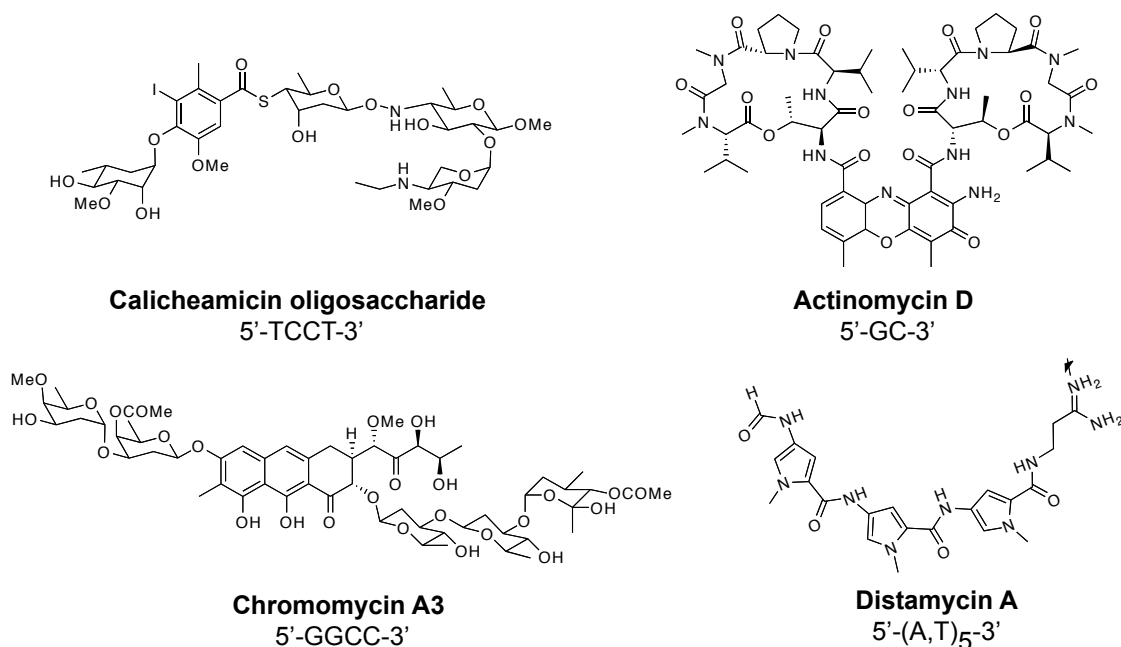


Figure 1.4 Chemical structures of DNA-binding small molecule natural products

activity has been attributed to interference of replication and transcription.¹⁸ Actinomycin D intercalates DNA preferentially at 5'-GC-3' sequences in a 1:1 ligand:DNA stoichiometry. Actinomycin D is known to inhibit transcription and potentially DNA replication and has been used as a chemotherapeutic as well.^{19,20} Distamycin A is an A,T-binding oligopeptide of three N-methylpyrrole (Py) carboxamide units. As can be seen in **Figure 1.5**, X-ray and NMR structural studies of Distamycin A reveal that this crescent-shaped molecule can bind DNA in either a 1:1 or 2:1 stoichiometry relative to DNA. The 2:1 complex forms in an antiparallel orientation and results in expansion of the minor groove relative to the 1:1 ligand-DNA complex.²¹⁻²³

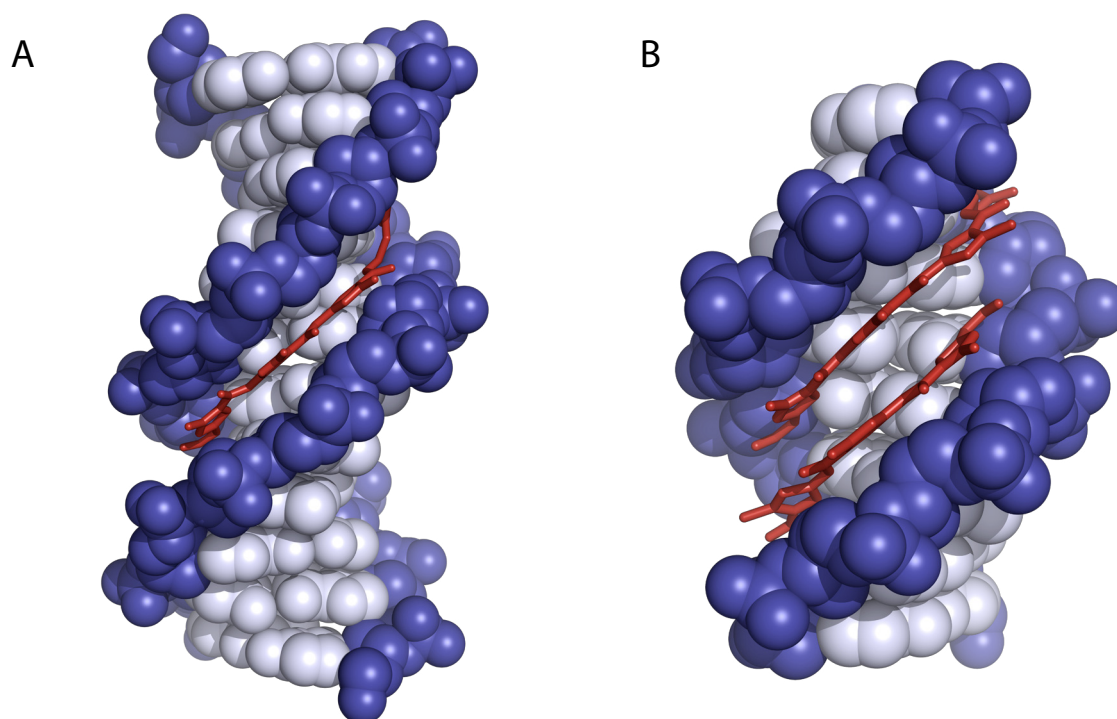


Figure 1.5 X-ray crystal structures of distamycin bound to DNA

A) Distamycin bound with a 1:1 ligand:DNA stoichiometry to the sequence 5'-CGCAAATTGCG' (PDB 2DND) B) Distamycin bound with a 2:1 stoichiometry to the sequence 5'GTATATAC-3'(PDB 378D)

The simple A,T binding natural product distamycin A has evolved over the past two decades into a new class of programmable heterocyclic oligomers that demonstrate high affinity and sequence specificity for the DNA minor groove.^{12,13} Incorporation of alternative heterocycles such as imidazole (Im) or hydroxypyrrole (Hp) expanded the sequence-recognition capabilities of polyamides and a set of pairing rules has been developed to allow for the programmable targeting of desired DNA sequences. Sequence-specific recognition of the minor groove of DNA by polyamides arises from the pairing of three different aromatic amino acids, pyrrole (Py), imidazole (Im), and hydroxypyrrole (Hp).²⁴⁻²⁶ The targeted binding surface of a crescent-shaped polyamide may be programmed by the incremental change of atoms on the corners of the ring pairs presented to the DNA minor groove floor. Stabilizing and, importantly, destabilizing interactions with the different edges of the four Watson-Crick bases are modulated by shape complementarity and specific hydrogen bonds.²⁷⁻²⁹ An Im/Py pair distinguishes G•C from C•G, T•A, and A•T, and likewise a Py/Im pair distinguishes C•G from G•C, T•A, and A•T. Im presents a lone pair of electrons to the DNA minor groove and can accept a hydrogen bond from the exocyclic amine of guanine.⁵ Additionally, the Hp/Py pair distinguishes T•A from A•T, G•C, and C•G.⁴⁻⁶ Hp projects an exocyclic OH group toward the minor groove floor that is sterically accommodated in the cleft of the T•A base pair, preferring to lie over T not A.²⁵

As can be seen in **Figure 1.6**, in addition to developing the pairing rules, another key step of the evolution of Im/Py polyamides is the covalent linkage of the two antiparallel heterocyclic strands by a gamma amino butyric acid (GABA) unit, forming a “hairpin” polyamide, demonstrating a 100-3600-fold increase in affinity relative to the unlinked homodimeric motif.^{30,31} Additionally, the incorporation of the turn linkage in the form of a GABA or substituted GABA turn allows the incorporation of unsymmetrical ring pairs for the targeting of non-palindromic DNA sequences.³²

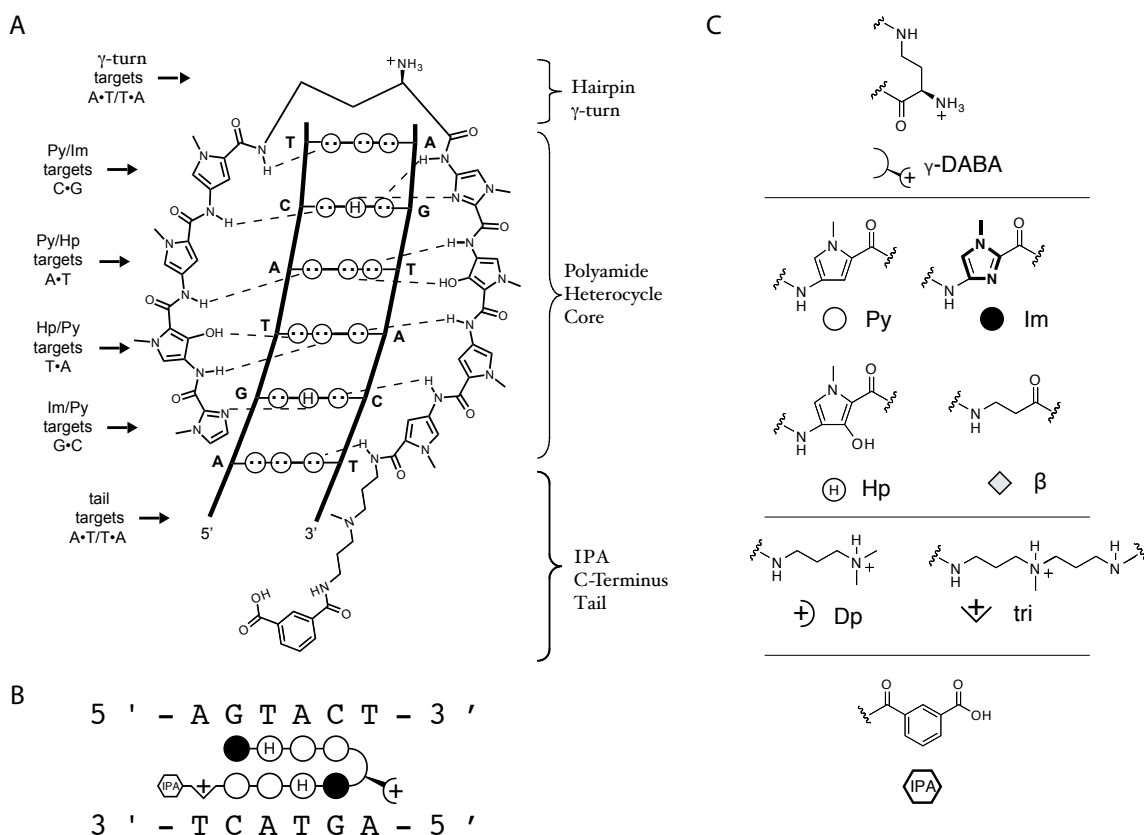


Figure 1.6 Recognition of the DNA minor groove by polyamides

A) Schematic illustration of the polyamide ImHpPyPy- γ -ImHpPyPy- β -Dp binding its target sequence, 5'-AGTACT-3'. Putative hydrogen bonds are shown as dashed lines. Hairpin turn and tail moieties, as indicated in illustration, tolerate T•A and A•T base pairs. B) A ball-and-stick representation of the polyamide shown in a. c) ball-and-stick symbols as well as abbreviations are listed below the chemical structures. This nomenclature will be used throughout this dissertation.

1.5 Polyamide inhibition of gene regulation in live cell culture

Given the fact that transcription is the key first step in gene expression, and the fact that the binding of *trans*-acting factors to promoter elements is critical for transcription, a key route to regulating gene expression lies in controlling the activity of these *trans*-acting factors. The programmability of Py-Im polyamides combined with the subnanomolar increases in affinity achieved by linking the two oligomeric strands provides polyamides with affinity competing with and often rivaling that of endogenous DNA-binding proteins.^{12,13} By using Py-Im polyamides to displace or prevent the binding of

transcription factors to their respective promoter DNA sequences, it is possible to modulate the expression of particular genes. Critical to this goal is the ability of polyamides to enter a live cell and permeate the nucleus in order to bind DNA and successfully regulate gene expression. Confocal microscopy studies have confirmed the positive nuclear uptake profiles of a variety of polyamide-fluorophore conjugates in a panel of cell lines.^{33,34} Additionally, the presence of an isophthalic acid (IPA) moiety in the tail region has been shown to yield high-affinity conjugates with improved nuclear permeability.³⁵

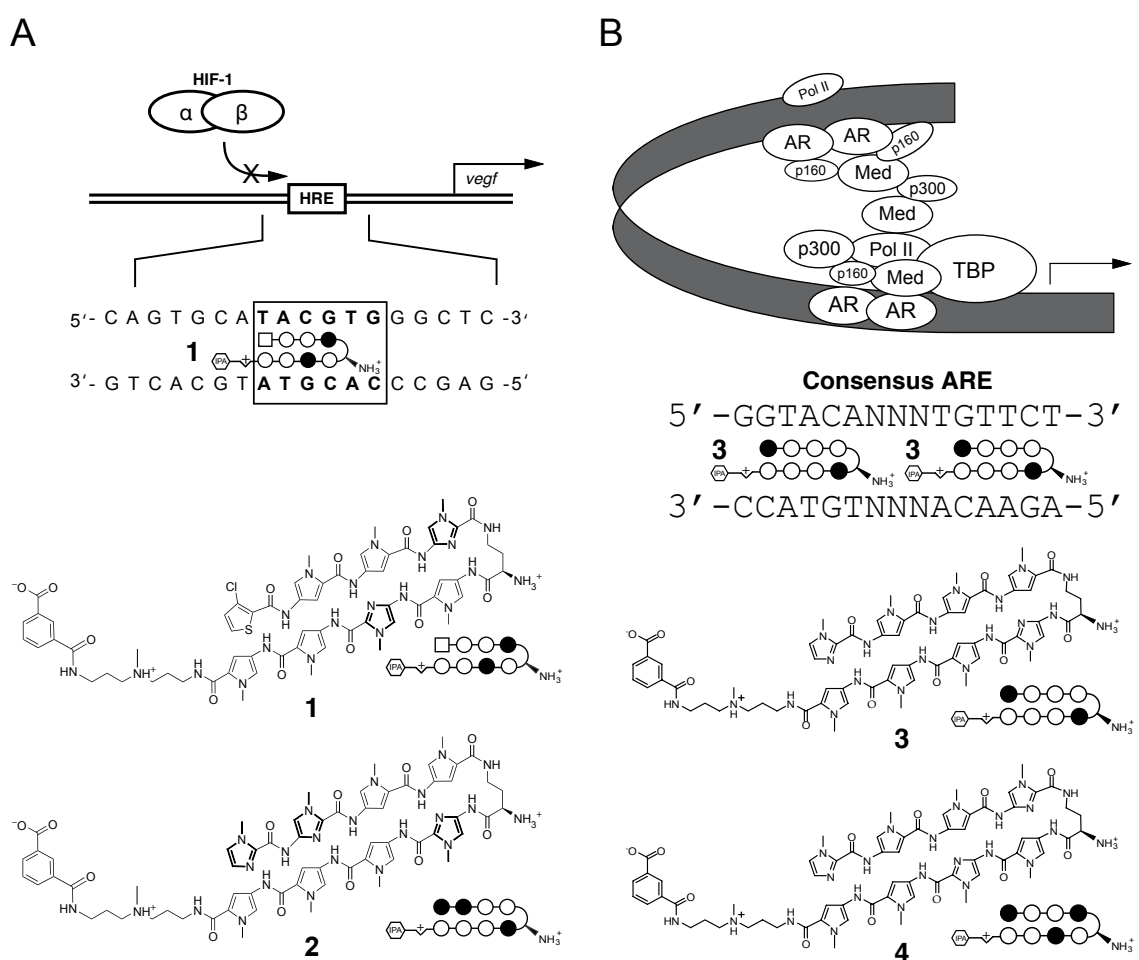


Figure 1.7 Polyamides as regulators of gene expression in cell culture

A) Schematic diagram of the VEGF promoter showing inhibition of HRE binding by HIF-1 (shown as HIF-1α/HIF-1β heterodimer), binding sequence of the HRE enhancer shown with match polyamide 1, and chemical structures and ball-and-stick models of match polyamide 1 and mismatch polyamide 2. B) Schematic diagram of the androgen receptor transcription complex, binding sequence of the consensus ARE targeted by match polyamide 3, and chemical structures and ball-and-stick models of match polyamide 3 and mismatch polyamide 4.

The culmination of the technological advances in this field has resulted in high affinity, sequence-specific, cell-permeable Py-Im hairpin polyamides that have successfully been utilized in several instances to modulate gene expression in cell culture. In a seminal example, a polyamide designed to bind to the hypoxia response element (HRE) was shown to disrupt the binding of hypoxia-inducible factor (HIF) to the HRE. Polyamide treatment was shown to decrease the transcription of vascular endothelial growth factor (VEGF) in cultured HeLa cells.^{35,36} In another example, a polyamide designed to target the androgen response element (ARE) has been shown to downregulate prostate-specific antigen (PSA) and other androgen responsive genes in prostate cancer cells.³⁷

The hypoxia inducible factor 1 α (HIF-1 α) transcription factor drives the expression of many genes in response to a low oxygen environment. HIF-1 α recognizes and binds a consensus sequence 5'-TACGTG-3' termed the Hypoxic Response Element (HRE). Polyamide **1** (**Figure 1.7A**) targets a subset of known HREs. One particular gene whose modulation is particularly interesting is vascular endothelial growth factor (VEGF), a gene responsible for the vascularization of tumors. When HeLa and U251 cells were treated with micromolar concentrations of polyamide **1**, a reduction in the deferoxamine (DFO) induction of VEGF gene was noted, while dosing with a non-HRE binding polyamide **2** did not produce a statistically significant change in gene expression. On a global scale, microarray experiments have shown that polyamide **1** downregulates a subset of genes upregulated by DFO induction. Furthermore, chromatin immunoprecipitation (ChIP) experiments demonstrated a reduced occupancy of HIF-1 α at the VEGF HRE in the presence of polyamide **1**.^{35,36,38}

The androgen receptor (AR) transcription factor binds as a homodimer to the androgen response element (ARE), 5'-GGTACAnnnTGTTCT-3', in response to induction by steroid hormones such as testosterone. AR-regulated gene expression is critical in the development and progression of prostate cancer. One key AR-regulated gene is prostate specific antigen (PSA),

a well-studied marker gene that correlates highly with the presence of prostate cancer. LnCap (prostate cancer) cells treated with a Py-Im hairpin polyamide **3** (**Figure 1.7B**) targeted to the consensus ARE half-site 5'-WGWWCW-3' demonstrate a suppressed dihydrotestosterone (DHT) induction of PSA mRNA transcription. This gene regulation modulation is not seen with treatment by a mismatch control polyamide targeting 5'-WGWCGW-3'. Microarray experiments assessing the global mRNA transcription in LnCap cells when treated with **3** and **4** indicate that **3** is able to disrupt induction by DHT for a subset of the DHT induced genes in a manner that is differential from treatment by **4** and likely is sequence-specific. ChIP experiments on **3** and **4** again suggest the disruption of a protein-DNA interface as a potential mechanism for polyamide activity.³⁷

1.6 An allosteric model for polyamide inhibition of steroid hormone receptors

Allosteric modulation of protein-DNA interactions

Both HIF-1 α and AR are major groove-binding proteins, yet minor groove-binding Py-Im polyamides are able to successfully disrupt protein DNA-binding and alter the gene regulation of the given transcription factor. A likely model for this inhibition is that allosteric modulation of the DNA occurs upon polyamide binding that renders the major groove of the DNA no longer capable of binding the transcription factor.^{39,40} The process of promoter recognition and utilization involves a stepwise interaction of a complex series of transcription factors with the promoter to create a stable DNA-protein complex that allows RNA polymerase to initiate transcription. There are as many as 100,000 protein-encoding genes in the mammalian genome, and rather than generate a unique transcription factor to regulate each gene, nature appears to have developed a limited number of transcription factors responsible for DNA recognition and that the high degree of specificity demonstrated is generated by specific protein-protein interactions that stabilize otherwise weak interactions on a promoter. Protein-DNA interactions are generally fairly

weak and will readily dissociate.⁴¹ The capacity to stabilize this otherwise weak interaction is likely a critical aspect of transcription control in which multiple factors must interact on the DNA to stabilize a functional complex.⁴

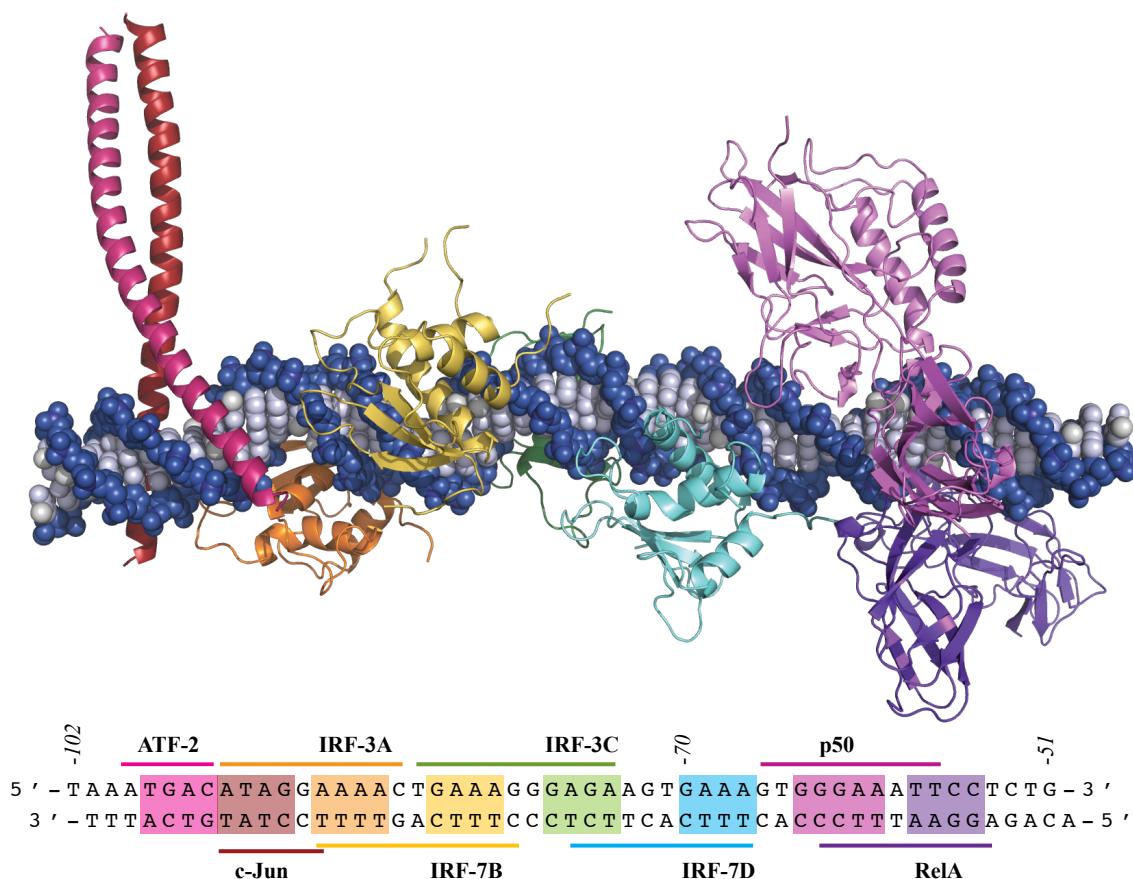


Figure 1.8 Atomic model of the cooperative assembly of interferon-β enhancesome

A composite model of allosterically driven protein–DNA recognition created from overlayed X-ray crystal structures (PDB 2O6G, 2O6I) showing 4–6 base pair transcription factor binding sites along the highly conserved composite DNA interface of 55 base pairs spanning approximately 160 Å in length.

Transcription factors often possess domains that are critical for mediation of protein-protein interactions with other components of the transcriptional machinery. In addition, transcription factors are also known to communicate indirectly through allosteric modulation of DNA resulting in cooperative assembly with very little direct protein-protein interaction. In this model, the sequence-specific binding of a transcription factor induces

perturbations in the DNA that modulate the binding of the next transcription factor. An elegant example of allosterically modulated protein–DNA specificity on a gene enhancer is the interferon-beta enhanceosome (**Figure 1.8**). In this protein complex, eight different transcription factors (ATF-2/c-Jun, IRF-3A, IRF-7B, IRF-3C, IRF-7D, p50, and RelA) cooperatively assemble on the enhancer, yet there are no protein-protein contacts between any of the proteins on this DNA sequence. Thus, it has been proposed that structural alterations to the DNA, such as widening or narrowing of the major or minor groove by individual protein–DNA interactions, create optimum binding shape and structure for other proteins, in a cooperative interaction.⁴² Each transcription factor binds four to eight base pairs and inhibition of the binding of any one of the proteins may result in interruption of the transcriptional activation activity of the protein complex as a whole.

Allosteric modulation of DNA by Py-Im polyamides

A recent high-resolution X-ray crystal structure of a β -amino turn-linked eight-ring cyclic Py-Im polyamide bound to the central six base pairs of the steroid hormone receptor consensus sequence reveals that significant modulation of DNA shape occurs upon polyamide binding. Cyclic polyamide **5**, comprised of two antiparallel ImPyPyPy strands capped by (*R*)- β -amino- γ turn units codes for the sequence 5'-WGWWCW-3'. The high-resolution crystal structure of the polyamide in complex with the 10 base pair DNA oligonucleotide sequence 5'-CCAGTACTGG-3', containing an ARE/GRE consensus DNA sequence demonstrates that structural alterations of DNA by these major groove-binding proteins and minor groove-binding cyclic polyamides operate in opposite directions.⁴⁰

Binding of cyclic polyamide **5** induces significant (>4 Å) widening of the DNA minor groove and compression of the major groove by more than 4 Å as compared to unliganded DNA (**Figure 1.9 C, D**).⁴³ Additionally, polyamide binding induces bending of

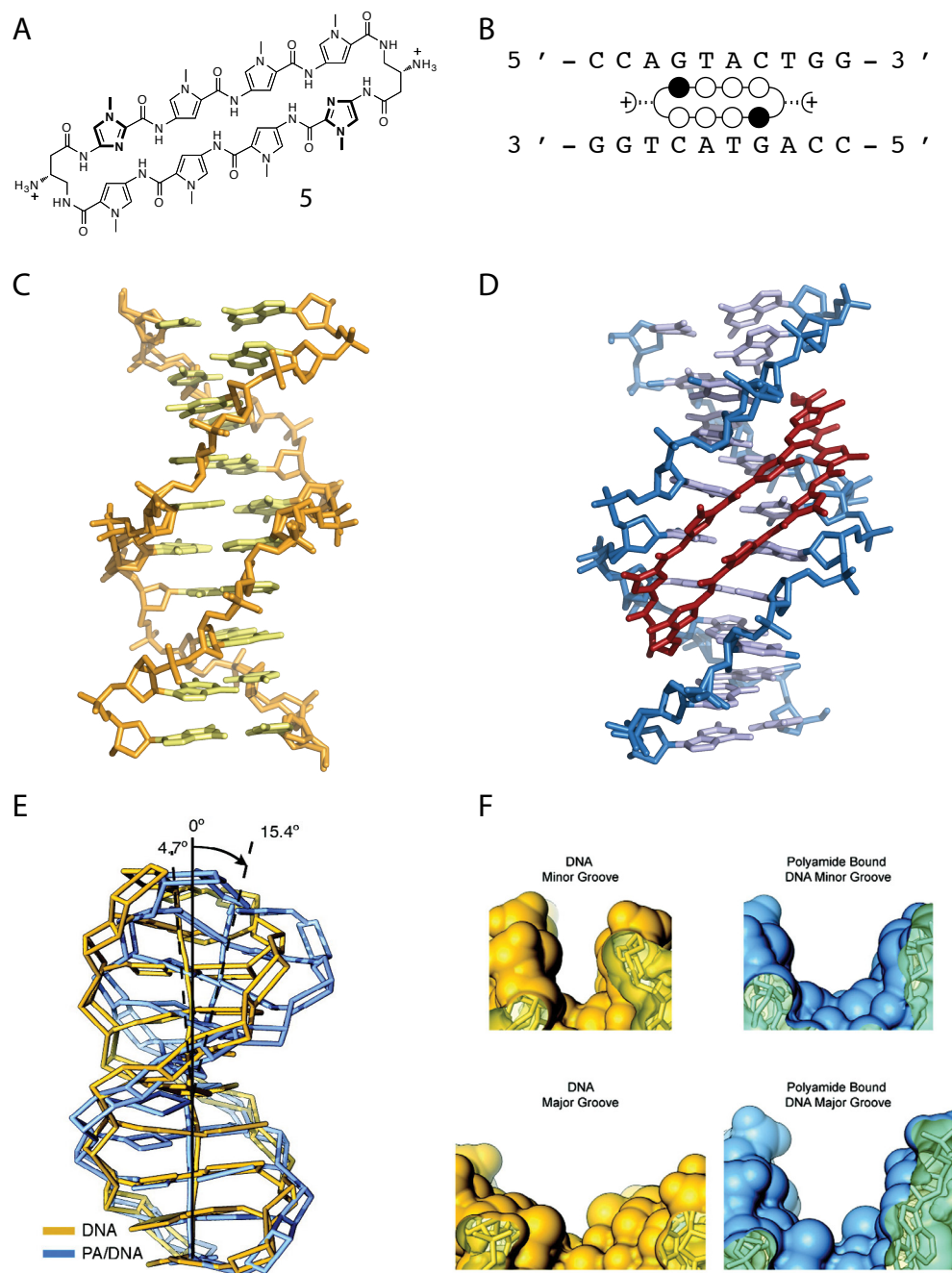


Figure 1.9 Comparison of native DNA to polyamide/DNA complex

A) chemical structure of cyclic polyamide 5 which targets 5'WGWWCW-3' *B)* ball-and-stick model superimposed over the binding site on the dsDNA oligonucleotide sequence used for crystallization. *C)* Native DNA crystal structure at 0.98 Å resolution (PDB 1D8G) *D)* DNA/polyamide co-crystal structure at 0.95 Å resolution. (PDB 3OMJ) *E)* Significant DNA bending is observed for polyamide-bound DNA (blue) versus unbound DNA (yellow). *F)* top: Comparison of the minor-groove width for DNA in the absence of polyamide (yellow) and in the presence of bound polyamide (blue). bottom: Comparison of the major-groove width for DNA in the absence of polyamide (yellow) and in the presence of bound polyamide (blue). In E and F, polyamide has been removed from the blue complex for clarity.^{39,40}

the DNA helix by $>15^\circ$ towards the major groove, resulting in major groove compression (**Figure 1.9E**). A slice through the short axis of the DNA helix, showing the minor and major groove geometry at the center of the polyamide binding site for uncomplexed and complexed DNA shows this distortion clearly (**Figure 1.9F**).⁴⁰

Ultimately, the perturbation in the major-groove geometry that occurs upon polyamide binding converts the wide, shallow surface of the major groove from a functionally exposed protein recognition domain to a narrow, deep cleft too small to accommodate the width of a standard protein α -helical domain or β -sheet from a transcription factor. A detailed analysis of the structure of DNA bound by the androgen receptor and the related glucocorticoid receptor reveals that the cyclic polyamide **5** is an allosteric modulator that perturbs the DNA structure in such a way that nuclear receptor protein binding is no longer compatible. This allosteric perturbation of the DNA helix provides a molecular basis for disruption of transcription factor–DNA interfaces by sequence-specific DNA binding polyamides.^{39,40}

1.7 Scope of this work

The work presented in this thesis is focused on the further development of Py-Im polyamides as tools for understanding the mechanisms behind gene regulation. By perturbing gene regulation in endogenous systems we can begin to understand the molecular basis for the difference between normal gene function and those alterations in gene control events that underlie certain disease states. In Chapter 2, we utilize a Py-Im polyamide targeted to bind 5'-WGWWCW-3' to probe the dual mechanism of transcriptional action displayed by the glucocorticoid receptor (GR) transcription factor. This sequence-specific small molecule probe is used to separate and identify a list of genes that are regulated by a protein-DNA interaction from those genes that are generally regulated by the GR in both DNA-binding-dependent and DNA-binding-independent mechanisms. In Chapter 3 we utilize polyamides to validate a new microarray-based tool for use in determining the DNA-sequence binding preferences of small molecules and proteins and in doing so examine the binding preferences of two polyamides that have been demonstrated to modulate gene expression in cell culture. The binding preferences for the polyamides assayed validate the polyamide pairing rules in an unbiased fashion. Chapter 4 presents progress towards divesting a group of genes regulated by the Myc-Max heterodimeric transcription factor. Our goal is to regulate genes bound at 5'-CACGTG-3' Myc-Max heterodimer binding sites while leaving Max-Max homodimer binding sites unbound. We anticipate that this will modulate the expression of Myc-regulated genes without inducing a Myc-upregulation feedback loop. In this chapter, a library of small molecules is developed to target the binding site, demonstrate that they bind with high affinity and specificity to the targeted site, and have positive uptake properties in a variety of live cells. While modulation of gene expression was not demonstrated in these cell lines, work is ongoing to probe this system further.

1.8 References

- (1) Crick, F. H. *Symposia of the Society for Experimental Biology* **1958**, 12, 138.
- (2) Lander, E. S.; Linton, L. M.; Birren, B.; Nusbaum, C.; Zody, M. C.; Baldwin, J.; Devon, K.; Dewar, K.; Doyle, M.; FitzHugh, W. *Nature* **2001**, 409, 860.
- (3) Venter, J. C.; Adams, M. D.; Myers, E. W.; Li, P. W.; Mural, R. J.; Sutton, G. G.; Smith, H. O.; Yandell, M.; Evans, C. A.; Holt, R. A. *Science* **2001**, 291, 1304.
- (4) Darnell, J. E. and Lodish, H. F. *Molecular cell biology*, 4th Ed.; W.H. Freeman: New York, 2000.
- (5) Williams, D. C.; Cai, M.; and Clore, G. M. *J. Biol. Chem.* **2004**, 279, 1449.
- (6) Juo, Z. S.; Chiu, T. K.; Leiberman, P. M.; Baikalov, I.; Berk, A. J.; Dickerson, R. E. *J. Mol. Biol.* **1996**, 261, 239.
- (7) Berkowitz, B.; Huang, D. B.; Chen-Park, F. E.; Sigler, P. B.; Ghosh, G. *J. Biol. Chem.* **2002**, 277, 24694.
- (8) Love, J. J.; Li, X.; Case, D. A.; Giese, K.; Grosschedl, R.; Wright, P. E. *Nature* **1995**, 376, 791.
- (9) Luisi, B. F.; Xu, W. X.; Otwinowski, Z.; Freedman, L. P.; Yamamoto, K. R.; Sigler, P. B. *Nature* **1991**, 352, 497.
- (10) Nair, S. K.; Burley, S. K. *Cell* **2003**, 112, 193.
- (11) Drew, H. R.; Wing, R. M.; Takano, T.; Broka, C.; Tanaka, S.; Itakura, K.; Dickerson, R. E. *Proc. Natl. Acad. Sci USA* **1981**, 78, 2179.
- (12) Dervan, P. B. *Bioorg. Med. Chem.* **2001**, 9, 2215.
- (13) Dervan, P. B.; Edelson, B. S. *Curr. Opin. Struct. Biol.* **2003**, 13, 284.
- (14) Rice, P. A.; Correll, C. C. *Protein-nucleic acid interactions: structural biology*; Royal Society of Chemistry, **2008**.

- (15) Waring, M. J. and Wakelin, L. P. G. In *DNA and RNA binders: from small molecules to drugs*; Wilson, W. D.; Bailly, C.; and Demeunynck, M., Editor.; Wiley-VCH Verlag GmbH, 2003; p. 1, 1-17.
- (16) Waring, M. J.; Britain, R. S. O. C. *Sequence-specific DNA binding agents*; RSC Publishing, 2006.
- (17) Bifulco, G.; Galeone, A.; Nicolaou, K. C.; Chazin, W. J.; Gomez-Paloma, L. *J. Am. Chem. Soc.* **1998**, *120*, 7183.
- (18) Hou, M. H.; Robinson, H.; Gao, Y. G.; Wang, A. H. *Nucleic Acids Res* **2002**, *30*, 4910.
- (19) Hou, M. H.; Robinson, H.; Gao, Y. G.; Wang, A. H. *Nucleic Acids Res* **2004**, *32*, 2214.
- (20) Kamitori, S.; Takusagawa, F. *J. Am. Chem. Soc.* **1994**, *116*, 4154.
- (21) Coll, M.; Frederick, C. A.; Wang, A. H.; Rich, A. *Proc. Natl. Acad. Sci USA* **1987**, *84*, 8385.
- (22) Mitra, S. N.; Wahl, M. C.; Sundaralingam, M. *Acta. Crystallogr. D. Biol. Crystallogr.* **1999**, *55*, 602.
- (23) Pelton, J. G.; Wemmer, D. E. *Proc. Natl. Acad. Sci USA* **1989**, *86*, 5723.
- (24) Kielkopf, C. L.; White, S.; Szewczyk, J. W.; Turner, J. M.; Baird, E. E.; Dervan, P. B.; Rees, D. C. *Science* **1998**, *282*, 111.
- (25) Urbach, A. R.; Szewczyk, J. W.; White, S.; Turner, J. M.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1999**, *121*, 11621.
- (26) White, S.; Szewczyk, J. W.; Turner, J. M.; Baird, E. E.; Dervan, P. B. *Nature* **1998**, *391*, 468.
- (27) Hays, F. A.; Teegarden, A.; Jones, Z. J.; Harms, M.; Raup, D.; Watson, J.; Cavaliere, E.; Ho, P. S. *Proc. Natl. Acad. Sci USA* **2005**, *102*, 7157.
- (28) Steitz, T. A. *Q Rev. Biophys.* **1990**, *23*, 205.

- (29) Turner, J. M.; Swalley, S. E.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1998**, *120*, 6219.
- (30) Herman, D. M.; Turner, J. M.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1999**, *121*, 1121.
- (31) Trauger, J. W.; Baird, E. E.; Dervan, P. B. *Nature* **1996**, *382*, 559.
- (32) Herman, D. M.; Baird, E. E.; and Dervan, P. B. *J. Am. Chem. Soc.* **1998**, *120*, 1382.
- (33) Edelson, B. S.; Best, T. P.; Olenyuk, B.; Nickols, N. G.; Doss, R. M.; Foister, S.; Heckel, A.; Dervan, P. B. *Nucleic Acids Res* **2004**, *32*, 2802.
- (34) Belitsky, J. M.; Nguyen, D. H.; Wurtz, N. R.; Dervan, P. B. *Bioorg. Med. Chem.* **2002**, *10*, 2767.
- (35) Nickols, N. G.; Jacobs, C. S.; Farkas, M. E.; Dervan, P. B. *Nucleic Acids Res* **2007**, *35*, 363.
- (36) Olenyuk, B. Z.; Zhang, G. J.; Kleo, J. M.; Nickols, N. G.; Kaelin, W. G.; Dervan, P. B. *Proc. Natl. Acad. Sci USA* **2004**, *101*, 16768.
- (37) Nickols, N. G.; Dervan, P. B. *Proc. Natl. Acad. Sci U S A* **2007**, *104*, 10418.
- (38) Nickols, N. G.; Jacobs, C. S.; Farkas, M. E.; Dervan, P. B. *ACS Chem. Biol.* **2007**, *2*, 561.
- (39) Chenoweth, D. M.; Dervan, P. B. *Proc. Natl. Acad. Sci USA* **2009**, *106*, 13175.
- (40) Chenoweth, D. M.; Dervan, P. B. *J. Am. Chem. Soc.* **2010**, *132*, 14521.
- (41) Stormo, G. D. Zhao, Y. *Nat Rev Genet* **2010**, *11*, 751.
- (42) Panne, D.; Maniatis, T.; Harrison, S. C. *Cell* **2007**, *129*, 1111.
- (43) Kielkopf, C. L.; Ding, S.; Kuhn, P.; Rees, D. C. *J. Mol. Biol.* **2000**, *296*, 787.

Chapter 2

Repression of DNA-Binding-Dependent Glucocorticoid Receptor-Mediated Gene Expression

The text of this chapter was taken in part from a manuscript co-authored with Nicholas G. Nickols and Peter B. Dervan (California Institute of Technology)

(Muzikar, K.A.; Nickols, N.G.; Dervan, P. B. “Repression of DNA-binding dependent glucocorticoid receptor-mediated gene expression” *Proc. Natl. Acad. Sci. USA.* **2009**, 106, 16598-16603)

Abstract

The glucocorticoid receptor (GR) affects the transcription of genes involved in diverse processes including energy metabolism and the immune response through DNA-binding-dependent and -independent mechanisms. The DNA-binding-dependent mechanism occurs by direct binding of GR to glucocorticoid response elements (GREs) at regulatory regions of target genes; the DNA-binding-independent mechanism involves binding of GR to transcription factors and co-activators that in turn contact DNA. A small molecule that competes with GR for binding to GREs could be expected to selectively affect the DNA-binding-dependent pathway by interfering with the protein–DNA interface. We show that a DNA-binding polyamide that targets the consensus GRE sequence binds the glucocorticoid induced zipper (*GILZ*) GRE, inhibits expression of *GILZ* and several other known GR target genes, and reduces GR occupancy at the *GILZ* promoter. Genome-wide expression analysis of the effects of this polyamide on a set of glucocorticoid induced and repressed genes could help elucidate the mechanism of GR regulation for these genes.

2.1 Introduction

Glucocorticoids are a major subclass of steroid hormones known to modulate a large number of metabolic, cardiovascular, immune, and behavioral functions and play diverse roles in growth, development, and maintenance of basal and stress-related homeostasis. Glucocorticoids exert a major influence on inflammatory processes and approximately 20% of the genes expressed in human leukocytes are regulated positively or negatively by glucocorticoids.¹ Glucocorticoids are among the most widely prescribed drugs worldwide, primarily used as anti-inflammatory and immunosuppressive agents.

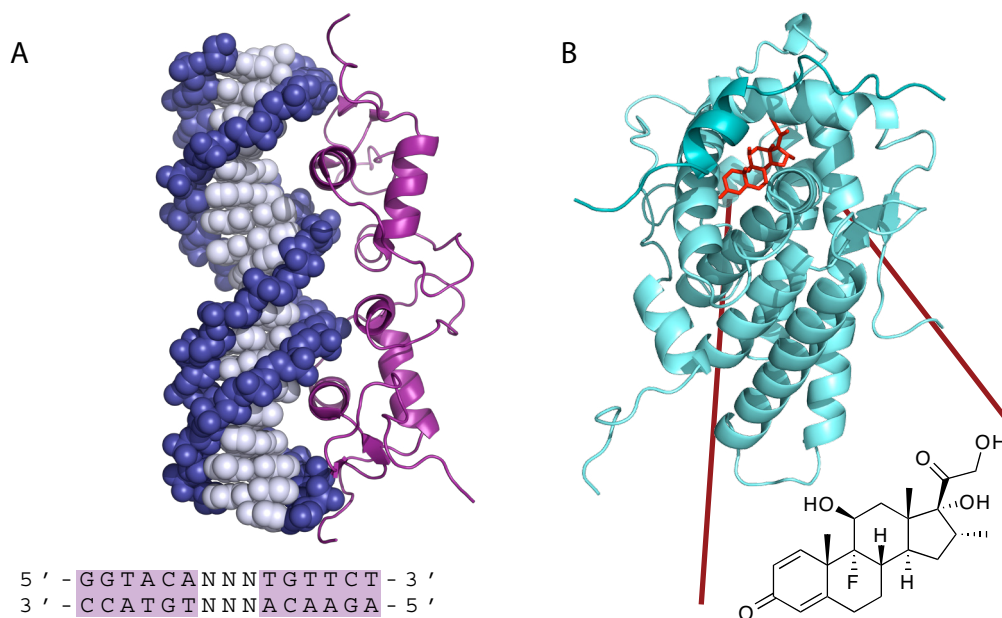


Figure 2.1 Crystallographic structures of the glucocorticoid receptor

A) DNA binding domain bound to DNA (PDB 1R4O) with consensus DNA binding site depicted below. B) Ligand binding domain bound to dexamethasone (PDB 1M2Z)

At the cellular level, the actions of glucocorticoids are mediated by a 94-kDa intracellular receptor protein, the glucocorticoid receptor (GR). The GR belongs to the superfamily of steroid/thyroid/retinoic acid receptor proteins that function as ligand-dependent transcription factors and binds with high affinity to glucocorticoids (GCs) such as cortisol

and dexamethasone. As a member of the same class of transcription factors, GR is structurally similar to the androgen and progesterone receptors, containing a zinc-finger motif DNA binding domain, a dimerization domain, and a ligand binding domain.² In the absence of ligand, GR resides in the cytoplasm of cells as part of a complex of proteins including chaperon heat shock proteins (hsps) 90, 70, and 50. Hsp90 regulates ligand binding, as well as cytoplasmic retention of GR by exposing the ligand binding site and masking the two nuclear localization sequences contained within the protein.³⁻⁵ Ligand binding releases GR from sequestration by cytoplasmic heat shock proteins⁵ and activates a series of cellular activities, which leads to nuclear localization and homodimerization (Figure 2.2).

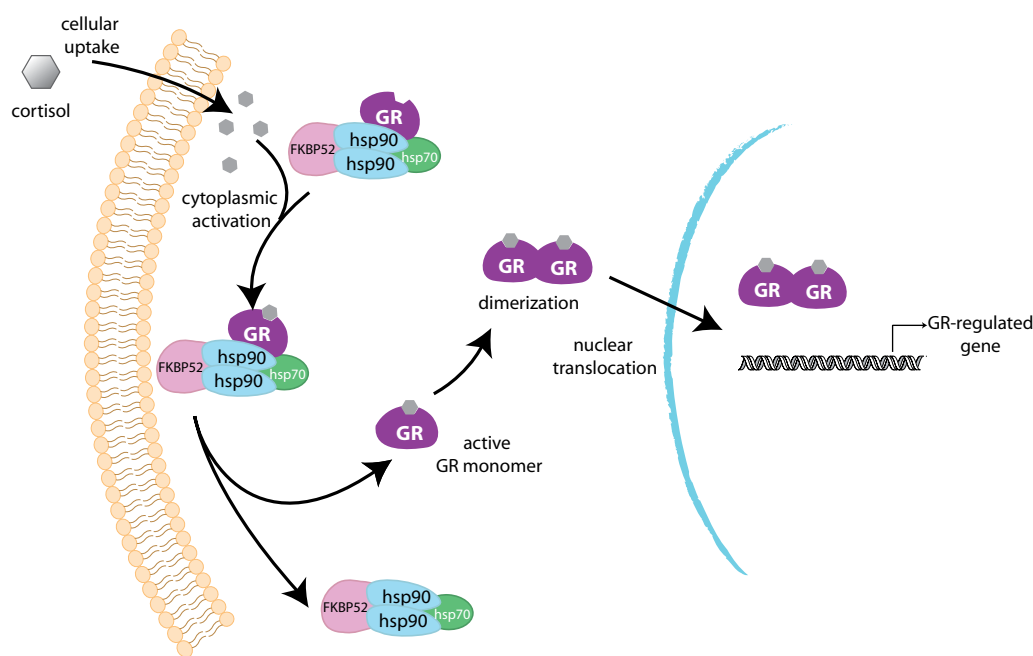


Figure 2.2 Simplified model for the response of glucocorticoid receptor to steroid hormone stimulus

The glucocorticoid hormone, cortisol, passes through the plasma membrane into the cytoplasm where it binds to the specific, high-affinity glucocorticoid receptor (GR), releasing it from sequestration by heat shock proteins. The activated GR forms a homodimer and is translocated to the nucleus where the receptor complex binds to specific DNA-responsive elements to activate gene transcription.

Like other steroid hormone receptors, GR is known to modulate gene transcription via the binding of receptor dimers to specific palindromic sequences called glucocorticoid response elements (GREs), usually located in the *cis*-regulatory region of target genes—this mode of action is termed transactivation. Additionally, the GR has been shown to exert its actions through an indirect, non-DNA-binding mechanism, termed transrepression, in which transcriptional modulation is achieved through crosstalk between GR and other transcription factors such as nuclear factor- κ B (NF- κ B),⁶ activator protein-1 (AP-1),^{7,8} Sma and Mad-related protein (Smad), and signal transduction and activator of transcription (STAT).⁹ This protein-protein cross talk does not require the DNA-binding activity of the GR, as GR mutants that are deficient in dimerization function have been shown to lose DNA binding ability as well as simple GRE-mediated transcription function but retain their transrepression activity.^{7,10}

Because many GR target genes are immune modulators, synthetic GR agonists such as dexamethasone are among the most effective anti-inflammatory drugs available for the treatment of a variety of chronic and acute inflammatory diseases. Unfortunately, because of the functions of other GR target genes, long term treatment with corticosteroids results in metabolic and behavioral derangements that can be treatment limiting. While the GR targets involved in inflammatory and immune regulation have not been comprehensively defined, there is a great deal of evidence suggesting that transrepression—that is, GR interaction with NF- κ B and/or AP-1 and the subsequent suppression of their target genes—is the major mechanism by which glucocorticoids achieve their desired anti-inflammatory effect.^{8,11,12}

An understanding of the mechanisms of GR activity on target genes has been explored using a variety of approaches including microarray analysis,^{13,14} ChIP-scanning,¹⁵ and modulation of GR activity using siRNA,¹⁶ genetic mutants,⁷ and ligands with modified

structures.¹⁷ However, these methods would not be expected to explicitly differentiate between the direct and indirect mechanisms of GR action. A small molecule that competes with GR for binding to the consensus GRE could be expected to specifically disrupt GR-DNA binding and be used as a tool to identify GR target genes whose regulation mechanism depends on a direct protein–DNA interface. This differentiation of mechanisms of GR target gene regulation could contribute to efforts to develop more specific GR modulators that retain immunosuppressant functions while minimizing side effects resulting from other GR targets.¹⁸ Py-Im polyamides have previously been used to modulate gene expression in cell culture via inhibition of the transcription factor–DNA interface of both hypoxia inducible factor^{19,20} and androgen receptor²¹ to their respective DNA response elements. Because Py-Im polyamides can be selectively programmed to recognize the known DNA-binding sequence of the GR, interruption of the GR–DNA binding interaction by polyamides represents a unique opportunity to inhibit the DNA-binding-dependent activity of endogenous GR while leaving the protein–protein-mediated activity unaffected (**Figure 2.3**).

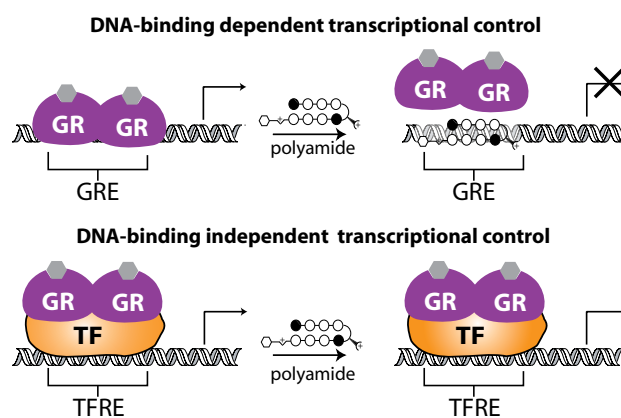


Figure 2.3 Effect of polyamide-DNA binding on GR gene regulation

Top: Direct mechanism is dependent upon GR-DNA binding at the GR response element (GRE). Bottom: Indirect mechanism is dependent upon GR binding to another protein, indicated here as a general transcription factor TF (with binding site TF response element TFRE). A sequence-specific polyamide designed to bind to the GRE but not to the TFRE would alter gene expression controlled by the DNA-binding-dependent but not the DNA-binding-independent mechanism.

In this study we designed a polyamide targeted to the sequence 5'-WGWWCW-3' (where W represents either a T•A or an A•T base pair), found in the consensus GRE, with the goal of disrupting GR-GRE binding (**Figure 2.3A**, polyamide **1**). This polyamide binds the two known GREs found in the promoter of the well-characterized glucocorticoid-induced leucine zipper (GILZ) gene, inhibits expression of GILZ and 17% of transcripts induced by dexamethasone in cultured alveolar epithelial cells (A549), and reduces GR occupancy at the GILZ promoter *in vivo*. A “mismatch” polyamide that targets the sequence 5'-WGWCGW-3' (**Figure 2.3A**, polyamide **2**) is used as a control for non-GRE binding polyamide effects. The subset of GR-regulated genes that are uniquely affected by the GRE-targeted polyamide may represent a set of genes that are regulated by GR through direct GR-GRE binding.

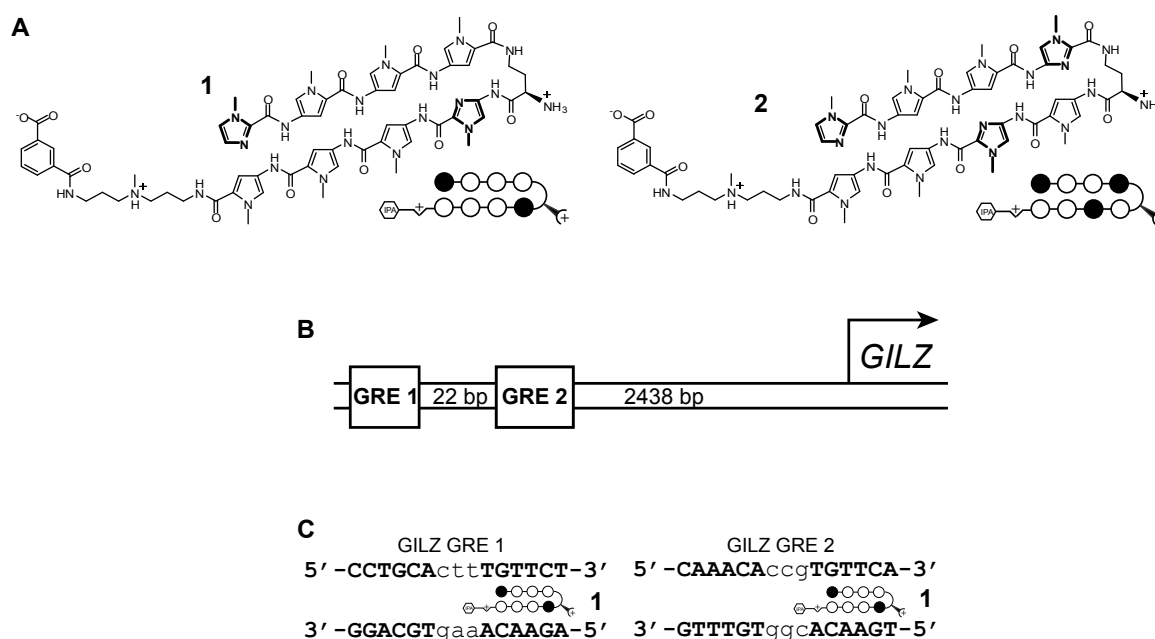


Figure 2.4 Polyamide design and *GILZ* promoter structure

A) Structure of match polyamide **1** designed to bind 5'-WGWWCW-3' and mismatch control polyamide **2** designed to bind 5'-WGWCGW-3', where W represents a T•A or an A•T basepair. Ball-and-stick models of polyamides represent the structure shown, imidazole and pyrrole monomer units are represented by filled and open circles, respectively. The isophthalic acid tail moiety is represented by a hexagon. B) Representation of the *GILZ* promoter region with its two functional GRE sites indicated. C) Sequences of the two *GILZ* GREs shown with polyamide **1** bound to its target site.

2.2 Results

I. Binding affinities of Py-Im polyamides to GRE1 and GRE2 of the GILZ promoter

The proximal GILZ promoter contains two functional GREs¹⁵ (GRE1: 5'-CCTGCActtTGTTCT-3' and GRE2: 5'-AAACAccgTGTTCA-3' spaced 22 base pairs apart approximately 2500 base pairs upstream of the transcription start site (**Figure 2.4 B,C**). The DNA binding affinity of polyamides **1** and **2** on this sequence was measured by quantitative DNase I footprint titrations using a 5' ³²P-labeled PCR fragment of pGR_GILZ, which contains a 78 base pair sequence from the promoter encompassing both functional GILZ GREs (**Figure 2.5**). Polyamide **1** has $K_a = 1.9 \pm 0.8 \times 10^{10} \text{ M}^{-1}$ for the GRE1 consensus half-site 5'-TGTTCT-3' and $K_a = 8.8 \pm 1.8 \times 10^9 \text{ M}^{-1}$ for the GRE2 consensus half-site 5'-TGTTCA-3'. Binding of polyamide **2**, which targets the sequence 5'-WGWCGW-3', to the GREs is not measurable by these methods ($K_a \leq 1 \times 10^7 \text{ M}^{-1}$). Binding of polyamide **2** with a K_a of $6.3 \pm 0.4 \times 10^9 \text{ M}^{-1}$ is observed at the site 5'-TGTCTT-3' located between the GREs. This is a single base pair mismatch site for polyamide **2**, and thus **2** can be expected to demonstrate some binding to this site.²²

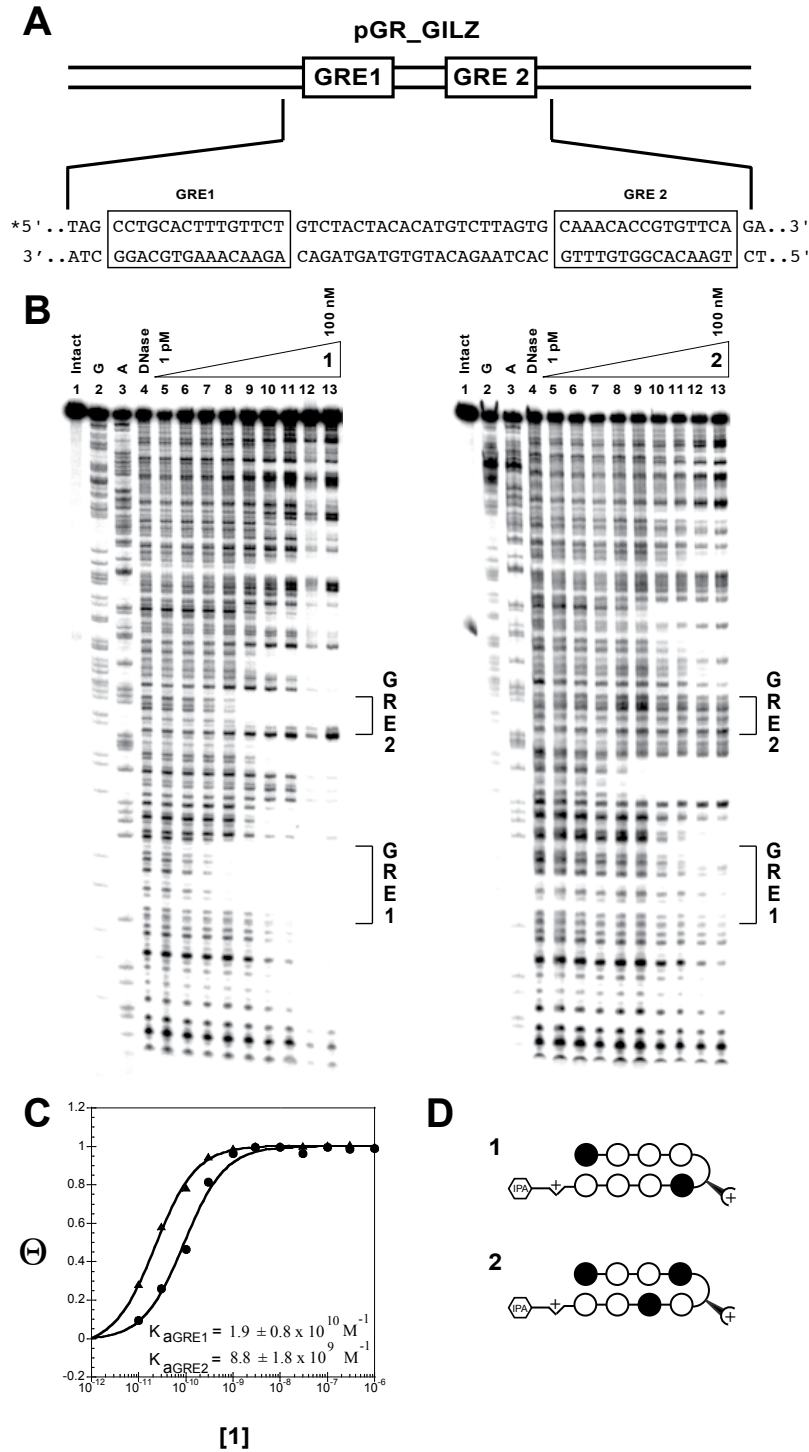


Figure 2.5 DNase I footprinting of *GILZ* promoter region

A) Sequence of the pGR_GILZ plasmid insert. *B)* Storage phosphor autoradiograms from quantitative DNase I footprint titrations of polyamides **1** and **2**. Lane 1: intact DNA, lane 2: G reaction, lane 3: A reaction, lane 4: DNase control, lanes 5-13: DNase I digestion products in the presence of 1 pM, 3 pM, 10 pM, 30 pM, 100 pM, 300 pM, 1 nM, 3 nM, 10 nM, 30 nM, 100 nM polyamide, respectively. *C)* binding isotherms for polyamides **1** and **2** bound to GRE **1**. *D)* Ball-and-stick models of polyamides **1** and **2**.

II. Electrophoretic mobility shift assay (EMSA)

The effects of polyamides **1** and **2** on the binding of recombinant human glucocorticoid receptor to an oligo containing the GRE1 site of the *GILZ* promoter were measured by an electrophoretic mobility shift assay (**Figure 2.6**). Incubation of the ^{32}P -labeled GRE DNA with recombinant human GR produces a gel shift that is reduced in the presence of polyamide **1** at concentrations as low as 10 nM. Polyamide **2** has minimal effect at the same concentrations. The shift is abolished by incubation with a 100-fold excess of unlabeled GRE DNA but is unaffected by similar treatment with a scrambled control DNA, indicating a specific shift resulting from a GR-GRE binding event.

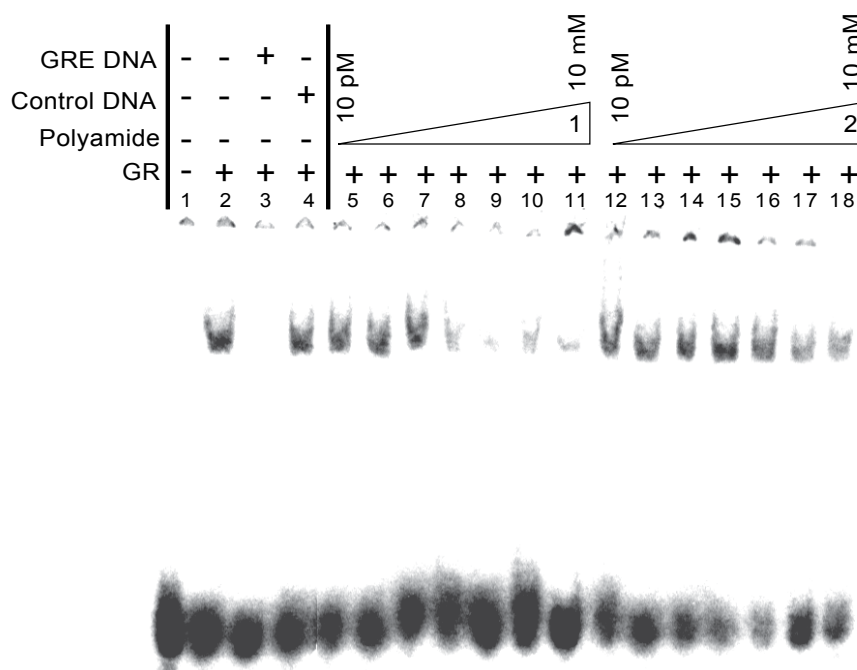


Figure 2.6 *In vitro* displacement of GR binding by polyamide **1**

Storage phosphor autoradiogram from EMSA of recombinant human glucocorticoid receptor binding to a 27-bp oligonucleotide duplex containing the *GILZ* GRE1. Lanes represent the following conditions, where * represents ^{32}P labeling: 1) Free *GRE DNA 2) *GRE DNA + GR 3) *GRE DNA + GR + GRE DNA 4) *GRE DNA + GR + scrambled DNA 5-11) *GRE DNA + GR + polyamide **1** (lanes 5-11) or **2** (lanes 12-18) in concentrations increasing from 10 pM, 100 pM, 1 nM, 10 nM, 100 nM, 1 μM to 10 μM , respectively.

III. Inhibition of glucocorticoid-induced GILZ expression

Induction of *GILZ* mRNA by dexamethasone in the presence of polyamides **1** and **2** in A549 cells was measured by quantitative real-time RT-PCR. Polyamide **1** inhibits the expression of dexamethasone-induced *GILZ* mRNA up to 65% at 5 and 10 μ M, as measured in this assay (**Figure 2.7A**). Polyamide **2** does not show a measurable effect on *GILZ* expression at these concentrations. The GR antagonist mifepristone was used as a control and inhibits the expression of *GILZ* up to 79% at 3 μ M. GR occupancy at the *GILZ* promoter was assessed by chromatin immunoprecipitation (**Figure 2.7B**). Six hour dexamethasone treatment results in a 15-fold increase in GR occupancy at the *GILZ* promoter, treatment of the cells with polyamide **1** for 48 hours prior to harvest reduces this occupancy, while treatment with the mismatch polyamide **2** shows a more modest effect. While polyamide **2** does not bind the GRE sites we cannot exclude the possibility that it may bind to other regions of the *GILZ* promoter. While polyamide **2** affects *GILZ* promoter occupancy, albeit significantly less so than **1**, it is somewhat surprising that **2** only minimally affects *GILZ* mRNA. However, it is not known what degree of occupancy is necessary for maximal induction of *GILZ* under these conditions. Treatment of cells by both polyamides **1** and **2** modestly affects cell proliferation and viability in a concentration and time dependent manner (**Figure 2.7C**).

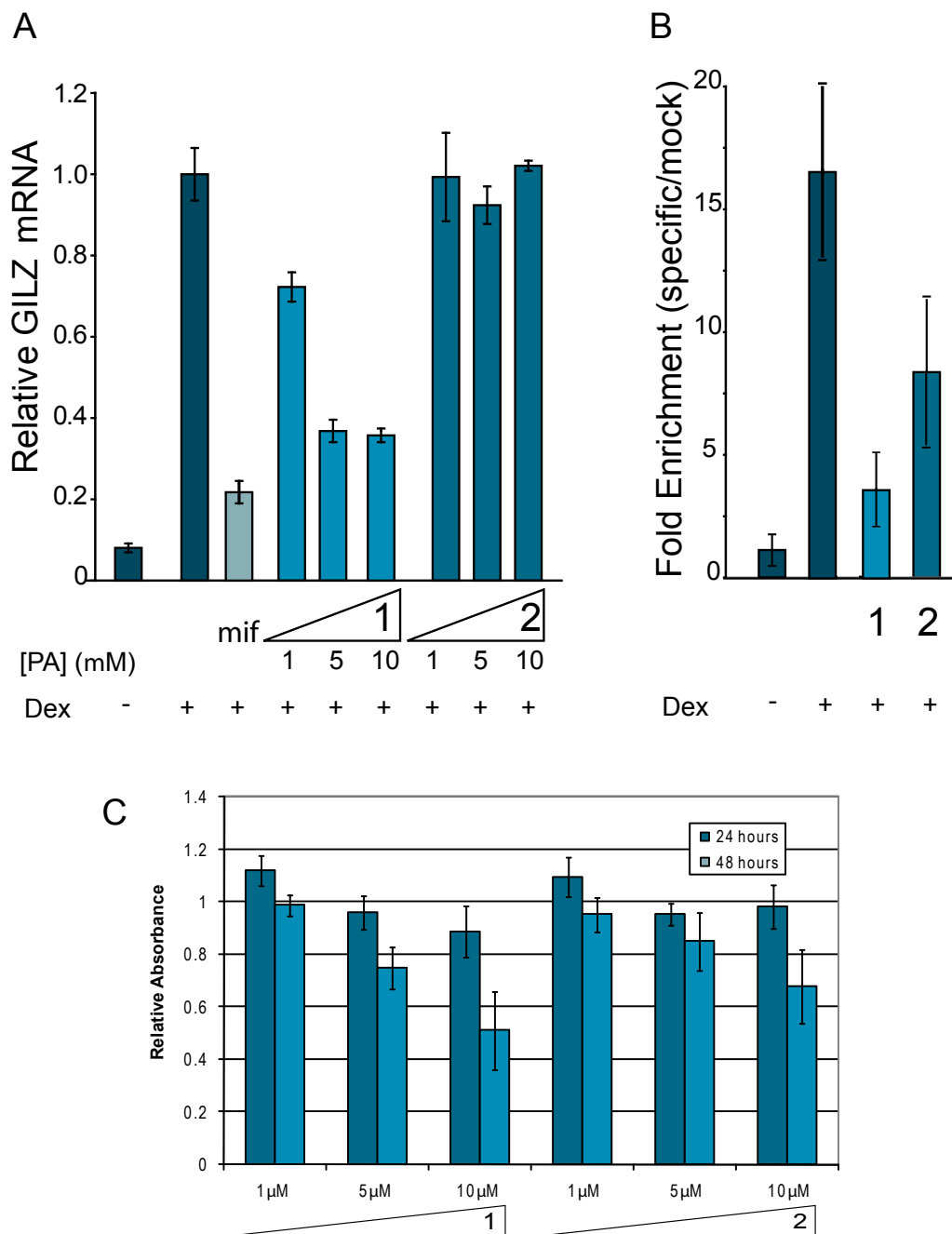


Figure 2.7 Inhibition of dexamethasone-induced *GILZ* expression by **1** and **2**

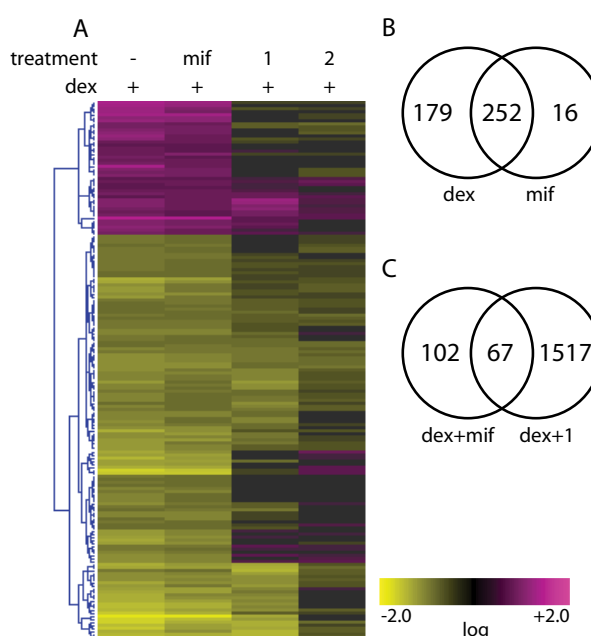
A) Induction of *GILZ* mRNA in the presence of mifepristone (mif), **1** and **2**, as measured by quantitative real-time PCR. Polyamide **1** inhibits expression of *GILZ* mRNA up to 65% at 5 and 10 μ M while polyamide **2** shows no effect. 3 μ M mifepristone control inhibits expression up to 79%. Error bars represent SD.

B) Chromatin immunoprecipitation assays with anti-GR or mock antibody treatment expressed as fold-enrichment (specific/mock) of DNA sequences at the *GILZ* promoter. GR occupancy at the *GILZ* promoter is decreased in the presence of **1** (10 μ M) and to a lesser extent by **2**. Error bars represent SD.

C) WST-1 colorimetric assay of cell viability and proliferation. Cells were treated for 24 and 48 hours with either polyamide **1** or **2** and subsequently assayed for cell viability and proliferation by measuring the absorbance of the dye at A_{max} of 450 nm on a multiwell plate reader.

IV. Genome-wide microarray analysis

Global effects of polyamides **1** and **2** and the GR antagonist mifepristone on gene expression in dexamethasone-stimulated A549 cells were monitored with Affymetrix high-density Human Genome U133 Plus 2.0 arrays, which interrogate >50,000 transcripts. Of these transcripts, 431 were affected greater than 2-fold by dexamethasone compared to non-induced control, with 323 transcripts induced while 108 were repressed. An agglomerative clustering analysis of the transcripts affected by dexamethasone demonstrates distinctive patterns of genes that are GC-responsive and GR-mediated but have differential response to treatment by polyamide **1** and **2** (**Figure 2.8A**) at 5 μ M concentration.



The GR antagonist mifepristone abolishes the activity of GR and therefore was used as a control to identify those transcripts that were affected by a GR-related mechanism. 431 transcripts were either induced or repressed by 2-fold or greater ($p < 0.01$) by dexamethasone treatment. Of these, the effects of dexamethasone were abolished by simultaneous treatment by 3 μ M mifepristone for 252 transcripts, indicating that these 252 were a result of GR activity (**Figure 2.8B**). For 76 transcripts of this set of 252, the effects of dexamethasone treatment were abolished by polyamide **1** to a greater extent than by polyamide **2**, indicating a sequence-specific polyamide effect on those transcripts (**Figure 2.8C**). This set of 76 transcripts corresponds to 67 genes, listed in **Table 2.1**.

Table 2.1 Genes affected >2-fold by dexamethasone and mifepristone whose activity is modulated by polyamide **1** and not by polyamide **2**

Gene	Fold Induction	Gene	Fold Induction	Gene	Fold Induction	Gene	Fold Repression
CDKN1C (N33167)	20	IGFBP1 (NM_000596)	4	49111_at (N80935)	3	NR4A2 (NM_006186)	6
FKBP5 (NM_004117)	14	CDC42EP3 (AI754416)	4	TIPARP (AL556438)	3	IER2 (NM_004907)	3
DNAJC15 (NM_013238)	14	PLEKHA7 (AA758861)	4	EPB41L4A (AU144565)	3	EREG (NM_001432)	3
FGD4 (AI949549)	11	CEP3 (AI801777)	4	CEBPD (NM_005195)	3	IER3 (NM_003897)	3
CIDEA (NM_022094)	8	LOC153346 (AU157049)	4	EMP1 (NM_001423)	3	NR0B1 (NM_000475)	3
EDN3 (NM_000114)	8	IL6R (NM_000565)	4	LOC54492 (AK026748)	3	MAFK (BG231691)	3
PTGER4 (AA897516)	8	ARRB1 (BE207758)	4	ARRB1 (BC003636)	3	CYP24A1 (NM_000782)	3
METTL7A (NM_014033)	7	CDC42EP3 (AL136842)	4	SOC1 (AB005043)	3	EDN1 (NM_001955)	2
FAM105A (NM_019018)	7	AKAP13 (NM_006738)	3	SCNN1G (AI985987)	3	NEIL3 (NM_018248)	2
ATAD4 (NM_024320)	6	244650_at (AA581439)	3	EPB41L4B (NM_019114)	3	PTGS2 (AY151286)	2
GOLSYN (NM_017786)	6	PKP2 (NM_004572)	3	ARRB1 (NM_004041)	3	RND1 (U69563)	2
CORO2A (AL515381)	5	FBXL16 (AI613010)	3	IL6R (NM_000565)	3		
TFCP2L1 (NM_014553)	5	CEBPD (AV655640)	3	MTIF (BF246115)	2		
RASSF4 (N49935)	5	FOXO3 (AV725666)	3	JPH2 (AA716165)	2		
CDH16 (NM_004062)	4	FLVCR2 (NM_017791)	3	SLC22A5 (NM_003060)	2		
ACSL1 (NM_021122)	4	MT2A (NM_005953)	3	RHOU (AL096776)	2		
MT1X (NM_002450)	4	KIAA0146 (AI363213)	3	MAN1C1 (NM_020379)	2		
AKAP13 (NM_006738)	4	43511_s_at (AI201594)	3	KIAA0232 (D86985)	2		
ACSL1 (NM_001995)	4	ETNK2 (NM_018208)	3				

These genes may represent glucocorticoid-responsive genes that are controlled via direct GR-binding mechanisms.

We anticipate that the regulation of these genes by GR involves a direct GR-GRE interaction. In order to analyze the potential GRE sequence-specific role of polyamide treatment within this subset of GR-regulated genes, those genes that showed similar effect from both polyamides **1** and **2** were considered to be non-GRE-specific effects and were

not included among those analyzed. Many of the genes identified by the microarray as dexamethasone-responsive are consistent with those previously identified in this cell line.¹⁵ A list of the effect of all treatment conditions on the GR-regulated genes is available in the Materials and Methods section of this chapter.

2.3 Discussion

A mechanistic understanding of cell signaling is fundamental to the development of improved medicines and diagnostics. Towards this end, the functions of individual biomolecules and their involved pathways have been explored using both biological and chemical methods. Genetic approaches include gene knockouts, dominant negative proteins, and siRNA to affect the activity of selected genes. Chemical approaches have utilized pre-existing, designed, or discovered small molecules that perturb a particular protein or specific protein-protein interaction, and have been employed successfully to characterize cell signaling pathways.²³⁻²⁵

The elucidation of transcription factor, coactivator, and corepressor activation, their gene targets, and mechanisms of activity is a focus of intense research interest. Genome-wide, high-throughput approaches including mRNA microarray analysis,²⁶ ChIP-chip,²⁷⁻²⁹ and more recently ChIP-sequencing³⁰ have been used to define the targets of transcription factors. Specific protein-DNA interactions are the interfaces where information from protein signaling is converted into programs of gene expression. Used in the context of genome-wide analysis, small molecules that perturb this interface in a predictable manner could become useful tools for understanding gene expression. Programmable, DNA-binding Py-Im polyamides offer a chemical approach to perturbing protein-DNA interactions that could be used for characterization of transcription factor-DNA interactions. We have reported previously that Py-Im polyamide **1**, which targets

5'-WGWWCW-3', inhibits androgen receptor (AR) binding to its consensus sequence 5'-GGTACAnnnTGTTCT-3'.²² As members of the same class of transcription factors, AR and GR share highly conserved DNA-binding domains, similar consensus sequences and transactivation mode of action. However, each acts within their individual biological contexts and responds to distinct chemical stimuli. The use of polyamide **1** in the context of disrupting the GR-GRE interface offers additional perspective in the mechanistic study of GR gene-regulatory action and its dual transactivation/transrepression mechanisms. Because of the central role of GR in multiple inflammatory response pathways and its role as a drug target, mechanistic studies of GR-modulated gene expression are a growing field of study. However, the multiple mechanisms by which GR mediates gene expression makes identification of GR target genes a challenge.^{9,18,31}

The complex transcriptional activity of the GR involves DNA-binding-dependent as well as DNA-binding-independent mechanisms that result in both gene activation as well as repression. In general the most well-understood mechanism is that by which genes are activated by direct DNA binding of GR to a GRE in the regulatory region of target genes followed by recruitment of coactivators and the general transcriptional machinery. Another well-documented, though less understood mechanism is the transrepression mechanism, in which GR exerts its influence indirectly by binding to other proteins and transcription factors leading to the repression of genes controlled by those proteins. Intriguingly, an increasing body of evidence indicates that much of the anti-inflammatory activity of GCs is mediated by the transrepression mechanism, a finding that has been attributed largely to the repression of key inflammatory transcription factors including AP-1 and NF- κ B.⁶⁻⁸ Meanwhile, many of the reported side effects of GC treatment have been attributed to transactivation mechanism.³² Efforts to develop new GC-based drugs that can dissociate anti-inflammatory and immunosuppressive effects from side effects have focused largely on establishing a method of dissociating the transrepression from the transactivation activity

of the GR. Much of the work in this area has centered about the development of new GR-binding ligands that display a more selective gene regulation pattern than classical GCs. The end result of this body of effort has demonstrated that the problem is complex, as slight alterations in ligand structure and chemical reactivity can have a pronounced influence on the transcriptional regulatory activity of GR.¹⁷

An additional complication to mechanistic study of the GR is the fact that examples have also been noted in which genes are repressed through DNA-binding-dependent GR action, as well as those in which genes are activated through DNA-binding-independent action. Techniques using immunoprecipitated chromatin fragments reveal a great deal of useful information about GR target sites, but are not expected to distinguish between sites of direct GR-GRE interaction and sites of indirect GR-protein-DNA interaction. Our approach utilizes a DNA binding polyamide targeting the consensus GRE that would be expected to dissociate the direct DNA-binding from the indirect, DNA-binding-independent GR gene-regulatory mechanisms.

In this study, a sequence-specific polyamide targeted towards the consensus DNA-binding sequence of the glucocorticoid receptor was tested on a well-known glucocorticoid-induced gene, *GILZ*, in order to establish its ability to disrupt GR-DNA binding and thereby regulate gene transcription. Polyamide **1**, designed to target the sequence 5'- WGWWCW-3', has been shown by DNase I footprinting to bind at subnanomolar concentrations to the right half-site of each of two different functional GREs located in the promoter region of the *GILZ* gene. Quantitative real-time PCR analysis of RNA isolated from A549 cells treated with polyamide **1** demonstrates a 60% reduction in dexamethasone-induced *GILZ* mRNA levels as compared with vehicle control. Chromatin immunoprecipitation indicates that in the presence of polyamide **1**, the dexamethasone-induced GR occupancy of the *GILZ* promoter is reduced, suggesting that it is polyamide

occupancy at this site that is responsible for the lowered mRNA expression levels. A control polyamide **2** targeted at a different sequence, 5'-WGWCGW-3', shows reduced affect on GR promoter occupancy and no effect on *GILZ* gene transcription.

The established ability of polyamide **1** to disrupt the GR-GRE interaction for *GILZ* led us to a genome-wide search for other transcriptional events that are interrupted sequence-specifically. Affymetrix microarrays interrogating >50,000 transcripts were chosen to examine the global effect of polyamides **1** and **2** at 5 μ M concentration on dexamethasone-treated cells. The GR antagonist mifepristone was used as a control in order to isolate which of the dexamethasone-responsive transcripts result from GR activity. These conditions isolated 252 genes which were considered genuine GR-modulated GC-affected transcripts. Both polyamides had a similar and modest effect on cell proliferation under these experimental conditions.

In order to tease out which of these effects is due to a binding event at a 5'-WGWWCW-3' sequence corresponding to a GRE, we have eliminated genes from the list that were affected by polyamide **2**, as both the footprinting and ChIP data shows that it is possible that this compound may have an effect on genes whose promoters contain a 5'-WGWCGW-3' site. Of the list of 252 GR-modulated transcripts, treatment with polyamide **1** had a unique effect on 170 transcripts over treatment with polyamide **2**. This left us with a final list of 170 transcripts whose expression is confidently affected both by GR as well as our sequence-specific small molecule. Of this final list of confidently interrogated transcripts we find 76 transcripts, corresponding to 67 genes, which are identified as genes whose expression was modulated by polyamide **1**. We believe this list of genes to represent 67 genes that are regulated by direct GR-DNA interactions. While little information exists in the literature linking glucocorticoid response with these genes, we note that several genes with known functional GRE binding sites are included in the list (**Table 2.2**). For genes

in which the binding site has been identified, the sequence is listed, with the predicted polyamide **1** target binding site underlined. The GREs for the genes listed in this table have been determined to have GC-induced GR occupation at the indicated location relative to the given gene. For some of these genes, functionality has been confirmed via point mutation or a luciferase-driven reporter assay. The fact that these genes contain polyamide **1** binding sites and demonstrate responsiveness to polyamide **1** treatment gives us confidence that the subset of genes identified here represents a new list of DNA-binding dependent genes to be explored by the GR field.

Table 2.2 Genes identified in our study that have previously been shown to have GR-occupied and/or functional GREs

Gene	GRE (5' → 3')	Location	Ref
CDKN1C	<u>AGAACAgcc</u> TGTCCT	Promoter	(32)
FKBP5	AGAACAggg <u>TGTTCT</u>	Intron	(31, 33)
ACSL1	AGCACAtcg <u>AGTTCA</u>	Intron	(31)
AKAP13	--	Exon	(31)
ETNK2	--	Intron	(31)
IGFBP1	--	Intron, Upstream	(31)
MT2A	AGGACAgccTGTCCT	Upstream	(31)
	AGAACAgga <u>TGTTTA</u>		

There is still very little known about many of the genes that are affected by glucocorticoids, including which genes are activated as a result of transrepressed mechanisms or transactivated mechanisms. The lack of identified GREs to be found in the literature complicates the search for genes that are regulated by the transactivation mechanism. DNA-binding polyamides represent a unique approach to differentiating the transrepression from transactivation activity of the GR—the potential to selectively block the protein-DNA interactions of endogenous GR while leaving the protein-protein interactions unaffected. This approach is limited by the target site degeneracy of both the polyamide as well as GR, effects of specific polyamide-DNA binding on programs of other

transcription factors, and non-specific effects of polyamides that are independent of sequence-specific polyamide-DNA binding. Small molecules that affect specific GR-protein interactions represent a complementary approach that is as yet unexplored. It is our hope that the list of genes provided by this study serves as a guide towards genes that necessitate further exploration in the search to selectively regulate the transactivation mechanism of glucocorticoid response.

2.4 Materials and Methods

Materials

Unless otherwise stated, DNA oligonucleotides were ordered PAGE-purified from Integrated DNA Technologies. Unless otherwise stated, reagents were purchased from Sigma-Aldrich. All solvents were purchased from Aldrich or EMD Biosciences. Trifluoroacetic acid (TFA) was purchased from Halocarbon. rac-Dithiothreitol (DTT) was purchased from ICN. RNase-free DEPC water was purchased from US Biochemicals. Water (18.2 M Ω) was purified using a Millipore water purification system.

Methods

UV spectra were recorded using an Agilent 8453 UV-Vis spectrophotometer. Polyamide concentrations were measured as a solution in water at $\lambda=310$ nm using an estimated extinction coefficient of $\epsilon=69,500$ M⁻¹ cm⁻¹ for 8-ring polyamides.³⁹ LASER desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) was performed using an Applied Biosystems Voyager DE Pro spectrometer. Analytical and preparative high-performance liquid chromatography (HPLC) were performed with a Beckman Gold system equipped with a diode array (analytical) or single-wavelength (preparative) detector.

i. Polyamide synthesis

Pyrole and imidazole monomer units [Boc-Py-OBt (1,2,3-Benzotriazol-1-yl 4-[(tert-Butoxycarbonyl)amino]-1-methylpyrrole-2-carboxylate) and Boc-Im-OH (4-[(tert-Butoxycarbonyl)amino]-1-methylimidazole-2-carboxylic acid) respectively] are synthesized according to established protocols³⁷ and maintained as general group stock. Fmoc-D(Dab)-Boc-OH turn monomer unit and Boc₂O were purchased from Peptides International.

Polyamides **1** and **2** were synthesized by solid-phase methods on Kaiser oxime resin (Nova Biochem, Darmstadt, Germany) following established protocols.³³ In a typical synthesis, 4 molar equivalents (relative to resin loading for the lot of oxime resin used) of monomer unit were activated with 4 equivalents of PyBOP (NovaBiochem) and 4 equivalents of diisopropylethylamine (DIEA) in a solution of dimethylformamide (DMF) (solution was approximately 0.3-0.5 M in monomer unit). In a glass peptide synthesis vessel, the resin was shaken with this solution for 2-4 hours at room temperature until the reaction was complete, as assessed by analytical HPLC of a cleaved sample of resin. Resin was deprotected by shaking in a solution of 20% trifluoroacetic acid (TFA) in methylene chloride for 20 minutes at room temperature. After completion of synthesis of the oligomer the Fmoc protecting group was removed by 10 minute incubation with 20% piperidine in DMF at room temperature. The free amine was subsequently protected by incubation of the resin for 1 hour in a solution of Boc₂O in DMF. The polyamide was subsequently cleaved from resin by incubation at 55 °C with 3,3-diamino-N-methyl-dipropylamine overnight (8-14 hours). The cleavage product was purified by reverse-phase HPLC, frozen in LN₂ and lyophilized to dryness.

Isophthalic acid (IPA) was activated with PyBOP (Nova Biochem) and conjugated to the polyamides as previously described.³⁴ Briefly: a 1 M solution of IPA in DMF and DIEA (50eq) was mixed with a 0.75 M solution of PyBOP in DMF. This solution was added to a lyophilized aliquot of cleaved polyamide and allowed to sit at room temperature for 1 hour. Following conjugation the polyamide was precipitated from solution by addition of cold diethyl ether and the precipitate was pelleted by centrifugation, the supernatant was removed and the pellet was allowed to air-dry. The Boc protecting group was removed by the addition of 100 μ L of neat TFA to the pellet. The deprotection proceeded for 10 min at room temperature before being diluted with 20% acetonitrile/water (0.1% TFA) followed by purification by reverse-phase HPLC. Purities and identities of the polyamides were assessed by HPLC, UV-visible spectroscopy, and MALDI-TOF MS.

1: (MALDI-TOF) $[M+H]^+$ calcd for $C_{65}H_{77}N_{22}O_{12}^+$ 1357.6, observed 1357.5

2: (MALDI-TOF) $[M+H]^+$ calcd for $C_{64}H_{77}N_{23}O_{12}^+$ 1358.6, observed 1358.6

ii. Plasmid preparation

Plasmid pGR_GILZ was constructed by ligating the following hybridized inserts (a 78-bp sequence from the GILZ promoter containing GRE1 and GRE2) into the BamHI / HindIII polycloning site in pUC19:

5'-GATC CCATA AGTATAGC CTGCACTTTG TTCTGTCT ACTACACATGT
CTTAGTGC AAACACCGTGT TCAGAGAG GTTGTG-3'

5'-AGCT CACAA CCTCTCTG AACACGGTGT TTGCACTA AGACATGTGTA
GTAGACAG AACAAAGTGCA GGCTATAC TTATGG-3'

The ligated plasmid was then transformed into JM109 subcompetent cells (Promega) by standard methods (30 minute incubation on ice followed by 45 second heat shock (42 °C) followed by 1 hour incubation at 37 °C). Colonies were selected for α -complementation on agar plates containing 50 mg/L ampicillin, 120 mg/L IPTG, and 40 mg/L X-gal after overnight growth at 37 °C. Cells were harvested after 16 h growth at 37 °C in LB medium containing 50 mg/L ampicillin. Plasmids were then purified by mini-prep kits. The presence of the desired inserts was determined by capillary electrophoresis dideoxy sequencing methods (Largen).

iii. Preparation of 5'-labeled DNA for DNase I footprinting

Two primer oligonucleotides, 5'-AATTCGAGCTCGGTACCCGGG-3' (forward, corresponding to EcoRI restriction enzyme cut site) and 5'-CTGGCACGACAGGTTTCCCGA-3' (reverse, corresponding to PvuII restriction enzyme cut) were constructed for PCR amplification. The forward primer was radiolabeled using [γ -³²P]-dATP (MP Biomedicals) and polynucleotide kinase (Roche) according to the manufacturer's protocol, followed by purification using ProbeQuant G-50 spin columns. The desired PCR product was generated from the plasmid pGR_GILZ using the primer pair and Expand High Fidelity PCR Core Kit (Roche) following the manufacturer's protocol. The labeled fragment was loaded onto a 7% nondenaturing preparatory polyacrylamide gel (5% cross-link), and the desired 283 base pair band was visualized by autoradiography and isolated. Chemical sequencing reactions were performed according to published protocols.³⁸

iv. Quantitative DNase I footprint titrations

All reactions were carried out in a volume of 400 μ L according to published protocols. Polyamides were equilibrated with the radiolabeled DNA for 14 h prior to DNase I cleavage at 23 °C. Quantitation by storage phosphor autoradiography and determination of equilibrium association constants were as previously described.³⁵

v. Electrophoretic mobility shift assay (EMSA)

The oligonucleotide 5'-GTATAGCCTGCACTTTGTTCTGTCTAC-3' representing a 27-bp section of the *GILZ* promoter containing GRE1 (underlined) was annealed to its complement and end-labeled with [γ -³²P]dATP and polynucleotide kinase (Roche) following the manufacturer's instructions. Aqueous solutions of polyamides **1** and **2** at the indicated concentrations were incubated with the duplex at room temperature for 2 h in 5 μ L 5x buffer containing 100 mM HEPES, pH 7.9, 300 mM KCl, 25 mM MgCl₂, 10 mM dithiothreitol (DTT), and 50% glycerol. Recombinant human glucocorticoid receptor (Affinity Bioreagents, 3.5 mg/ml) was diluted 1:20 with 50 μ g/ μ L bovine serum albumin (BSA) and stored as a frozen stock solution. Fresh aliquots of this solution were used for each experiment and discarded after use. 2 μ L of the diluted GR protein, 1 μ L of 200 ng/ μ L poly dI-dC, 1 μ L of unlabeled DNA (for control lanes) were taken up in water to a final volume of 15 μ L and incubated at 4 °C for 30 min. This solution was added to the polyamide-DNA complex and incubated at room temperature for 1 h before being separated on a 6% polyacrylamide gel in 1x TBE. Gels were visualized on a phosphorimager.

vi. Measurement of dexamethasone-induced mRNA.

RNA isolation:

A549 cells (ATCC) were plated in 24-well plates at a density of $20\text{--}25 \times 10^3$ cells per well ($40\text{--}50 \times 10^3$ cells per mL) in F12-K medium (ATCC) supplemented with 10% FBS (Irvine Scientific, Santa Ana, CA) and 4 mM penicillin/streptomycin. After 24 h, the medium was replaced with F12-K containing 10% charcoal-stripped FBS, 4 mM penicillin/streptomycin, and polyamides or mifepristone at the designated concentrations. Cells were grown for an additional 48 h and then treated with 100 nM dexamethasone for 6 h. (**Figure 2.9**). The medium was removed, cells were washed with ice-cold PBS and immediately lysed with RLT buffer from an RNeasy kit (Qiagen). Further RNA isolation was carried out with the RNeasy kit as described in the manufacturer's manual. The isolated total RNA was quantified. The yields were 12-15 μg per well.

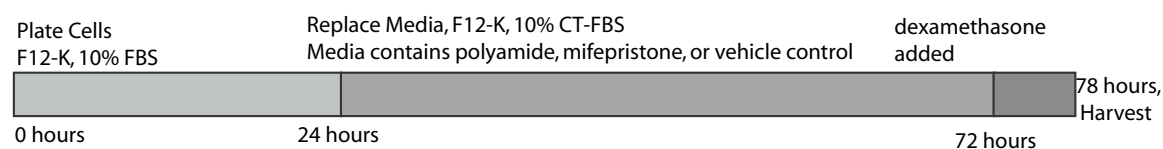


Figure 2.9 Timeline of cell treatment protocol, used for RT-PCR, ChIP and microarray assays

Reverse transcription:

A 2.5 μg sample of total RNA was used to reverse transcribe cDNA using Superscript II reverse transcriptase (Invitrogen) according to the manufacturer's protocol. Random hexamers were used as primers. The total volume for each RT reaction was 20 μL .

Real-time quantitative RT-PCR:

Analysis was performed using the GILZ gene primers described below, purchased from Integrated DNA Technologies. Quantitative real-time RT-PCR was performed with SYBR

Green PCR Master Mix (Applied Biosystems, Foster City, CA) following the manufacturer's suggested protocol. Temperature cycling and detection of the SYBR Green emission were performed with an ABI 7300 real-time instrument using Applied Biosystems Sequence Detection System version 1.2. Statistical analysis was performed on three independent experiments. mRNA of the genes of interest were measured relative to β -glucuronidase as an endogenous control. Primer sequences were designed using Primer3.³⁶

To amplify the 97-bp fragment from the 3'-translated region of GILZ:

Forward primer: 5'- CTCCCCGTTTGTTTTTCTCA -3'

Reverse primer: 5'- TGCTCCTTCAGGATCTCCAC -3'

To amplify the β -glucuronidase gene as an endogenous control:

Forward primer: 5'- CTCATT TGGAATTTTGCCGATT -3'

Reverse primer: 5'- CCGAGTGAAGATCCCCTTTTTA -3'

vii. Measurement of Cell Proliferation and Viability.

A549 cells were plated in 96-well plates, 100 μ L at 20×10^3 cells/mL in F12-K medium containing 10% charcoal-stripped FBS and 4 mM penicillin/streptomycin. After the indicated time period, 10 μ L of Cell Proliferation Reagent WST-1 (Roche, Cat. No. 05 015 944 001) was added to each well. Reagent was allowed to incubate at 37 °C for 30 minutes and absorbance was read at 450 nm on a Perkin-Elmer Victor 3 multiwell plate reader.

viii. Chromatin Immunoprecipitation.

A549 cells were plated in 15 cm diameter plates at a density of 10×10^5 cells per plate. Media conditions, polyamide treatment, time course, and dexamethasone stimulation were identical to the conditions described above for qPCR. Upon completion of the 6 h dexamethasone treatment, cells were cross-linked by treatment with 1% formaldehyde for 10 minutes. Chromatin was isolated using a ChIP-IT kit (Amersham) following the manufacturer's protocol. Chromatin was sheared and immunoprecipitated by overnight incubation at 4 °C with anti-GR antibody N499 (A gift from Keith Yamamoto, UCSF). A 1:1 mixture of Protein G and Protein A Agarose beads (Upstate) was used to isolate the immunoprecipitated material via centrifugation. Cross-links were reversed and the DNA was isolated via phenol/chloroform extraction followed by ethanol precipitation. qPCR using primers targeted to the GILZ promoter were used to assess enrichment of bound fragments as compared with mock-precipitated (no antibody) controls. PCRs were monitored with SYBR Green PCR Master Mix (Applied Biosystems) on an ABI 7300 qPCR instrument. Primer sequences, described below, were designed using Primer3.³⁶ Detailed procedure can be found in Nicholas Nickols' PhD dissertation.⁴⁰

To amplify the 88-bp fragment from the 5'-promoter region of GILZ:

Forward primer: 5'- GCACTGATTCATGGGTACTGG -3'

Reverse primer: 5'- ACCAACTCAGGACCAAAGGAG -3'

ix. Analysis of Gene Expression with Oligonucleotide Microarrays.

A549 cells were plated in 12-well plates at a density of $40\text{-}50 \times 10^3$ cells per well. Media conditions, polyamide treatment, time course, and dexamethasone stimulation were identical to the conditions described above for qPCR and ChIP. Mifepristone was added

at the same time as polyamide was added. RNA was isolated as previously described. The RNA was submitted to the Millard and Muriel Jacobs Gene Expression Facility at the California Institute of Technology where labeled mRNA was hybridized to Affymetrix high-density Human Genome U133 Plus 2.0 arrays according to established protocols. Gene expression was analyzed by using Resolver (Rosetta Biosoftware, Seattle, WA). Data were uploaded to the Gene Expression Omnibus repository (accession no. GSE17307). **Table 2.3** lists the effect of all treatment conditions on the GR-regulated genes.

2.5 Acknowledgements

We thank Keith Yamamoto for his kind gift of GR antibody. This work was supported by National Institutes of Health Grant GM051747. Mass spectrometry analyses were performed in the Spectrometry Laboratory of the Division of Chemistry and Chemical Engineering at the California Institute of Technology, supported in part by the National Science Foundation Materials Research Science and Engineering program. Oligonucleotide microarray experiments were performed in the Millard and Muriel Jacobs Genetics and Genomics Laboratory at the California Institute of Technology.

Table 2.3 Microarray fold changes under all treatment conditions of sequences affected ≥ 2 -fold by dexamethasone and mifepristone

Sequence	Accession #	Dex	Mif	1	2	Sequence	Accession #	Dex	Mif	1	2
CDKN1C	N33167	20.3	-18.9	-5.0	-1.1	STOM	M81635	3.5	-2.8	-1.8	-1.5
FKBP5	NM_004117	14.3	-5.0	-5.9	-2.9	AKAP13	NM_006738	3.5	-2.6	-4.1	-3.1
RRAD	NM_004165	13.8	-9.6	-1.4	2.0	NEXN	AF114264	3.5	-2.9	-4.4	-3.2
DNAJC15	NM_013238	13.7	-12.1	-15.4	-2.4	244650_at	AA581439	3.5	-3.1	-2.2	-1.1
TFCP2L1	A1928242	12.9	-5.5	-5.5	-2.0	PKP2	NM_004572	3.4	-3.3	-3.4	-1.5
CDKN1C	NM_000076	12.6	-7.8	-3.6	1.0	THBD	NM_000361	3.4	-2.3	-1.3	-1.9
FKBP5	W86302	12.1	-5.2	-5.8	-2.1	ANKRD44	AU157224	3.4	-3.1	-5.8	-5.5
RRAD	NM_004165	11.4	-11.4	-1.3	1.8	FOXO3	NM_001455	3.4	-2.5	-2.4	-2.1
FKBP5	A1753747	11.2	-4.5	-4.9	-1.8	FGD4	AA296351	3.4	-3.1	-2.6	-2.1
TSC22D3	AL110191	10.9	-6.4	-1.8	-1.3	FOXO3	N25732	3.4	-2.8	-3.3	-2.1
FGD4	A1949549	10.8	-6.1	-7.5	-2.0	LOC283278	A1690465	3.4	-2.6	-5.9	-2.5
CIDEC	NM_022094	8.4	-4.7	-5.4	-4.0	KIAA1462	AL553774	3.4	-3.4	-1.9	-2.4
CDKN1C	N95363	8.3	-5.8	-2.6	1.3	FBXL16	A1613010	3.4	-3.7	-2.3	-1.2
EDN3	NM_000114	8.2	-4.3	-5.8	-1.3	CEBPD	AV655640	3.3	-3.1	-4.4	-2.3
CDKN1C	R78668	7.9	-5.7	-3.3	1.1	FOXO3	AV725666	3.3	-2.9	-3.4	-1.9
PTGER4	AA897516	7.6	-4.3	-3.1	-1.9	SNAI2	AI572079	3.3	-3.1	-2.0	-2.0
FGD4	A1277617	7.6	-5.8	-6.9	-1.6	BIRC3	U37546	3.2	-2.8	1.9	1.4
FAM43A	AW264102	7.2	-3.9	-1.2	1.5	CALD1	AL577531	3.2	-2.5	-3.0	-2.3
METTL7A	NM_014033	6.6	-3.2	-7.2	-1.8	TSC22D3	NM_004089	3.2	-3.0	-1.6	-1.3
UNC5B	AK022859	6.6	-4.8	1.1	2.0	227082_at	A1760356	3.2	-2.1	-4.1	-3.7
FAM105A	NM_019018	6.5	-4.6	-2.8	-1.2	ALPP	NM_001632	3.2	-2.7	-1.2	1.7
ANGPTL4	NM_016109	6.3	-3.4	-1.6	-1.5	ANPEP	NM_001150	3.2	-2.6	-1.9	1.4
ATAD4	NM_024320	6.2	-2.8	-2.9	-1.0	ZBTB20	AW499525	3.2	-2.1	-8.1	-4.5
GOLSYN	NM_017786	5.9	-3.6	-5.1	-1.6	CALD1	NM_018495	3.2	-2.8	-3.1	-2.0
RGS2	NM_002923	5.5	-4.2	-1.4	-1.6	RBM20	AI539118	3.2	-3.1	-4.5	-3.8
CORO2A	AL515381	5.4	-3.4	-3.1	1.2	MT1P2	AF333388	3.2	-2.8	-2.4	-2.0
TFCP2L1	NM_014553	5.2	-3.5	-3.2	-1.5	MT1F	M10943	3.2	-2.8	-2.5	-1.9
ANGPTL4	AF169312	5.1	-3.0	-1.5	-1.6	FLVCR2	NM_017791	3.1	-2.9	-2.3	-1.4
CDKN1C	D64137	5.1	-4.2	-2.5	1.2	SERPINE1	NM_000602	3.1	-2.3	-7.7	-6.2
GMPR	NM_006877	5.0	-3.1	-1.1	2.1	213158_at	AA045174	3.1	-2.2	-21.5	-5.7
RASSF4	N49935	4.9	-4.4	-6.0	-2.0	MT1H	NM_005951	3.1	-2.9	-2.4	-1.9
PER1	NM_002616	4.9	-2.6	1.3	1.1	MT2A	NM_005953	3.1	-2.9	-2.3	-2.0
LOC286167	AV721528	4.5	-3.9	-1.2	1.0	KIAA0146	AI363213	3.1	-2.7	-4.1	-1.8
ZFP36	NM_003407	4.5	-3.2	1.0	-1.3	ABHD2	AI557319	3.1	-2.2	-3.6	-3.1
CDH16	NM_004062	4.4	-4.3	-2.6	1.4	GPR115	W67511	3.0	-2.5	-4.6	-4.3
ACSL1	NM_021122	4.3	-2.8	-4.5	-2.0	FOXO3	BE888885	3.0	-2.5	-2.7	-1.8

Accession numbers given are Affymetrix accession numbers.

Table continued on page 70

Table 2.3 Microarray fold changes under all treatment conditions of sequences affected ≥ 2 -fold by dexamethasone and mifepristone (*continued from page 69*)

Sequence	Accession #	Dex	Mif	1	2	Sequence	Accession #	Dex	Mif	1	2
C13orf15	NM_014059	4.3	-4.4	-2.0	1.2	CHST7	NM_019886	3.0	-2.7	1.7	1.9
MT1X	NM_002450	4.3	-3.9	-3.3	-2.2	43511_s_at	AI201594	3.0	-2.2	-2.9	-1.1
GRAMD4	AB018310	4.2	-4.3	-1.1	1.4	213156_at	BG251521	3.0	-2.0	-10.6	-4.8
SDPR	BF982174	4.2	-3.4	1.4	1.3	227121_at	BF476076	3.0	-2.2	-6.8	-4.0
AKAP13	M90360	4.2	-2.5	-5.9	-3.3	EPB41L4A	NM_022140	3.0	-2.1	-3.2	-2.2
ACSL1	NM_001995	4.1	-2.9	-4.2	-2.0	AKAP13	AW575773	2.9	-2.4	-4.4	-2.4
KCNB1	L02840	4.0	-4.0	-4.5	-4.3	EPB41L4B	AB032179	2.9	-2.4	-2.2	-2.1
AKAP13	AK022014	4.0	-2.3	-4.8	-3.1	SLC4A11	AF336127	2.9	-2.8	-1.4	1.1
DUSP1	AA530892	3.9	-2.2	-1.2	-1.4	SH3TC1	NM_018986	2.9	-2.7	-1.6	-1.0
SDPR	NM_004657	3.9	-2.9	1.4	1.2	ERRFI1	AL034417	2.9	-2.0	1.2	1.0
MT1X	NM_005952	3.9	-3.5	-2.8	-2.2	CALD1	NM_004342	2.9	-2.4	-3.6	-2.2
IGFBP1	NM_000596	3.8	-2.6	-2.5	-1.6	ETNK2	NM_018208	2.9	-2.6	-2.6	-1.5
ABHD2	AI832249	3.8	-2.3	-4.1	-3.7	PALM2-AKAP2	NM_007203	2.9	-2.2	-2.0	-2.5
CDC42EP3	AI754416	3.7	-2.2	-3.6	-1.9	MOBK2B	AI692878	2.9	-2.7	-6.0	-5.1
PLEKHA7	AA758861	3.7	-3.0	-5.8	-2.1	49111_at	N80935	2.9	-2.1	-2.7	1.0
STOM	AI537887	3.7	-2.7	-1.8	-1.5	BAIAP2	BC002495	2.8	-2.5	-1.2	1.1
CEP3	AI801777	3.7	-2.3	-3.7	-1.6	GADD45A	NM_001924	2.8	-2.6	1.2	1.3
REEP1	BE535746	3.6	-3.0	-6.6	-3.8	TIPARP	AL556438	2.8	-2.2	-2.3	-1.6
LOC153346	AU157049	3.6	-3.6	-2.2	-1.3	EPB41L4A	AU144565	2.8	-2.4	-3.2	-1.5
THBD	NM_000361	3.6	-2.3	-1.3	-1.8	CEBPD	NM_005195	2.8	-2.3	-2.1	-1.4
IL6R	NM_000565	3.6	-2.6	-2.9	-2.1	PRKCD	NM_006254	2.8	-2.1	-1.1	1.2
RAB11FIP1	NM_025151	3.6	-2.9	-1.2	1.3	MAOA	AA923354	2.8	-2.0	-3.3	-2.0
THBD	AW119113	3.6	-2.6	-1.5	-1.9	EMP1	NM_001423	2.8	-2.4	-2.7	-1.5
ABHD2	BE671816	3.6	-2.3	-4.2	-3.3	LOC54492	AK026748	2.8	-2.7	-2.2	-1.0
ARRB1	BE207758	3.5	-2.6	-4.5	-2.2	ARRB1	BC003636	2.7	-2.2	-3.3	-1.7
CDC42EP3	AL136842	3.5	-2.2	-3.3	-2.0	SOCS1	AB005043	2.7	-2.7	-2.0	1.1
BAIAP2	AB017120	2.7	-2.7	-1.4	-1.0	RHOU	AB051826	2.2	-2.1	-1.1	-1.0
EPB41L4B	NM_019114	2.7	-2.3	-2.3	-1.9	DNER	BF059512	2.1	-2.3	-3.4	-2.6
EMP1	BF445047	2.7	-2.7	-2.7	-2.3	COBLL1	NM_014900	2.1	-2.1	-3.0	-2.1
KIAA1462	AL050154	2.7	-3.3	-2.3	-2.5	BAIAP2	NM_017450	2.1	-2.6	-1.3	1.2
ARRB1	NM_004041	2.7	-2.0	-2.7	-1.7	HSD11B2	NM_000196	2.1	-2.0	-1.3	1.6
EMP1	NM_001423	2.7	-2.3	-2.7	-1.5	KIAA1545	AA527531	2.1	-2.2	-1.8	-1.8
B3GNT5	BE672260	2.7	-2.1	1.1	1.1	MAN1C1	NM_020379	2.1	-2.1	-2.0	-1.4
TFCP2L1	AW195353	2.7	-2.2	-2.3	-1.8	PPL	NM_002705	2.1	-2.3	-1.9	-1.4
SERPINE1	AL574210	2.6	-2.1	-3.9	-6.0	FLVCR2	AW001026	2.1	-2.1	-1.6	-1.3
NEXN	NM_144573	2.6	-2.4	-3.8	-3.0	KIAA0232	D86985	2.1	-2.1	-3.1	-1.4

Accession numbers given are Affymetrix accession numbers.

Table continued on page 71

Table 2.3 Microarray fold changes under all treatment conditions of sequences affected ≥ 2 -fold by dexamethasone and mifepristone (*continued from page 70*)

Sequence	Accession #	Dex	Mif	1	2	Sequence	Accession #	Dex	Mif	1	2
IL6R	AV700030	2.6	-2.2	-2.2	-1.5	AKAP2	BE879367	2.1	-2.1	-1.9	-2.2
RAB11FIP1	AA143793	2.6	-2.8	-1.1	1.3	GPR37	U87460	-2.0	2.3	1.5	1.8
TNS4	AA158731	2.6	-2.2	-1.1	-1.2	FOXA2	AB028021	-2.0	2.0	1.5	1.3
NDRG1	NM_006096	2.6	-2.4	-1.6	1.2	FZD7	AI333651	-2.0	2.1	3.3	2.1
PLEKHA2	BF347859	2.5	-2.1	-2.8	-2.6	LHX8	BC040321	-2.0	2.0	-1.1	-1.2
ITGB4	AF011375	2.5	-2.1	-2.1	-2.0	VASH2	AI961235	-2.1	2.1	1.5	1.6
STARD13	AA128023	2.5	-2.4	-2.9	-2.9	EDN1	NM_001955	-2.1	2.1	2.3	1.8
KIAA1462	AL553774	2.5	-2.4	-1.6	-1.9	KCNK3	NM_002246	-2.1	2.9	-1.4	-1.3
LIFR	AA701657	2.5	-2.2	-2.7	-3.0	SLC40A1	AL136944	-2.2	2.1	-1.4	1.1
RHOB	BI668074	2.5	-2.1	-1.1	-1.7	FAM113B	BF056901	-2.2	2.0	1.4	-1.1
CORO2A	NM_003389	2.5	-2.1	-1.8	1.0	TGFA	NM_003236	-2.2	2.2	1.8	2.1
CKB	NM_001823	2.5	-2.5	-1.4	-1.1	MYOCD	AI093327	-2.3	2.0	-1.0	-1.0
SEC14L2	NM_012429	2.5	-2.4	-2.1	-2.1	TGFB2	NM_003238	-2.3	2.3	-1.1	-1.1
CTGF	M92934	2.5	-2.9	-2.9	-2.2	LOC285513	AK026379	-2.3	2.3	1.8	1.3
FOS	BC004490	2.5	-2.1	1.1	-1.7	NEIL3	NM_018248	-2.4	2.0	2.7	1.6
MT1F	BF246115	2.4	-2.4	-2.2	-1.4	PTGS2	AY151286	-2.4	2.2	3.3	1.8
MT1G	NM_005950	2.4	-2.2	-1.8	-1.6	EDN1	J05008	-2.5	2.1	2.7	2.0
SRGN	J03223	2.4	-2.0	-1.3	-1.3	RND1	U69563	-2.5	2.3	2.3	1.5
PER2	NM_022817	2.4	-2.1	1.4	1.4	240173_at	AI732969	-2.5	2.3	-1.1	-1.2
THBS1	NM_003246	2.4	-2.2	-4.4	-2.7	SLITRK6	AI680986	-2.6	2.7	-1.7	-1.7
KIAA0232	AF143884	2.4	-2.2	-7.4	-1.9	GREM1	AF154054	-2.6	2.3	-1.9	-2.0
JPH2	AA716165	2.4	-2.5	-2.5	-1.9	RBM24	AI677701	-2.6	2.1	1.8	1.4
PACSIN2	BC008037	2.4	-2.1	-2.0	-2.3	PTGS2	NM_000963	-2.7	2.5	3.7	2.0
THBS1	AV726673	2.4	-2.2	-3.8	-2.5	HMOX1	NM_002133	-2.7	2.4	1.2	-1.6
ZBTB20	AW974823	2.4	-2.4	-1.5	-1.1	FAM84B	AL039862	-2.7	2.0	8.6	3.6
ARRB1	BF446943	2.4	-2.1	-2.4	-1.8	LONRF2	AV709727	-2.8	2.0	1.3	-1.1
PACSIN2	NM_007229	2.4	-2.0	-2.1	-2.1	IER2	NM_004907	-2.8	2.2	3.1	1.5
AHNAK	BG287862	2.3	-2.1	-1.5	-1.4	HLCS	AI682088	-2.8	2.1	1.4	1.1
HPCAL1	NM_002149	2.3	-2.2	-3.2	-3.0	SLC7A11	AB040875	-2.9	2.4	1.2	-1.6
213817_at	AL049435	2.3	-2.2	-1.9	1.3	EREG	NM_001432	-2.9	2.3	4.8	1.6
KLF6	BU683415	2.3	-2.1	-1.2	-1.4	IER3	NM_003897	-3.0	2.3	5.4	3.5
KLF6	BG250721	2.3	-2.0	-1.2	-1.3	GREM1	NM_013372	-3.0	2.5	-1.0	-1.5
LIFR	AI680541	2.3	-2.3	-2.8	-2.4	SLC7A11	AA488687	-3.1	2.7	1.2	-1.5
EMP2	NM_001424	2.3	-2.2	-1.2	1.3	NR0B1	NM_000475	-3.2	2.6	2.2	1.2
LRRC8A	AK024649	2.3	-2.1	-1.8	-2.1	LOC644943	AA876179	-3.2	2.8	1.8	1.2
SLC19A2	AF153330	2.3	-2.1	-1.2	-1.1	MAFK	BG231691	-3.2	2.2	4.0	1.7

Accession numbers given are Affymetrix accession numbers.

Table continued on page 72

Table 2.3 Microarray fold changes under all treatment conditions of sequences affected ≥ 2 -fold by dexamethasone and mifepristone (*continued from page 71*)

Sequence	Accession #	Dex	Mif	1	2	Sequence	Accession #	Dex	Mif	1	2
SLC22A5	NM_003060	2.3	-2.2	-2.1	1.0	CYP24A1	NM_000782	-3.3	3.0	2.1	1.0
KLF4	BF514079	2.3	-2.6	1.9	1.4	DIO2	AI038059	-3.3	2.7	1.2	-1.7
S100P	NM_005980	2.3	-2.1	1.1	-1.0	ID4	AL022726	-3.4	2.4	-1.5	-1.4
MOBK2B	AI375115	2.3	-2.2	-3.3	-3.5	TGFB2	M19154	-3.6	3.3	1.0	-1.3
TMEM43	W74580	2.3	-2.2	-1.4	-1.3	NAV3	NM_014903	-3.8	2.1	-1.0	-1.2
MT1E	AL031602	2.2	-2.3	-1.8	-1.6	NR4A2	AI935096	-3.9	2.5	40.9	5.8
ITGB4	NM_000213	2.2	-2.1	-1.5	-1.5	TM4SF20	NM_024795	-4.2	3.6	-1.0	-1.2
EPB41L4A	NM_022140	2.2	-2.1	-1.9	-1.6	SOX2	AI669815	-4.3	2.7	-1.2	-1.1
KLHL29	BE465475	2.2	-2.3	-4.9	-4.1	MYOCD	AI452798	-4.6	3.1	-1.5	-1.3
GPR115	NM_153838	2.2	-2.0	-3.1	-2.8	ST8SIA4	AA552969	-5.3	3.8	-1.2	-1.2
RHOU	AL096776	2.2	-2.4	-2.5	1.3	ST8SIA4	AA352113	-5.6	5.1	-1.2	-1.4
RHOB	AI263909	2.2	-2.4	-1.1	-1.5	NR4A2	NM_006186	-5.7	3.2	53.5	7.0
AKAP2	BG540494	2.2	-2.3	-2.1	-2.0	ST8SIA4	AI422986	-5.9	4.6	-1.3	-1.3
ENTPD2	AW134837	2.2	-2.3	-1.7	-1.5	IL11	NM_000641	-5.9	2.3	-1.1	1.1
KCNG1	AI332979	2.2	-2.1	-1.5	1.1	NR4A2	S77154	-6.0	3.1	63.6	7.0
TMEM43	AA115485	2.2	-2.1	-1.4	-1.3	GDF15	AF003934	-8.4	4.9	16.2	5.7
ST3GAL1	NM_003033	2.2	-2.2	1.0	1.2	229242_at	BF439063	-9.2	7.6	2.1	1.9

Accession numbers given are Affymetrix accession numbers.

2.6 References

- (1) Galon, J.; Franchimont, D.; Hiroi, N.; Frey, G.; Boettner, A.; Ehrhart-Bornstein, M.; O'Shea, J. J.; Chrousos, G. P.; Bornstein, S. R. *FASEB J* **2002**, *16*, 61.
- (2) Giguere, V.; Hollenberg, S. M.; Rosenfeld, M. G.; Evans, R. M. *Cell* **1986**, *46*, 645.
- (3) Bamberger, C. M.; Schulte, H. M.; Chrousos, G. P. *Endocr. Rev.* **1996**, *17*, 245.
- (4) Pratt, W. B. *J. Biol. Chem.* **1993**, *268*, 21455.
- (5) Pratt, W. B.; Toft, D. O. *Endocr. Rev.* **1997**, *18*, 306.
- (6) McKay, L. I.; Cidlowski, J. A. *Mol. Endocrinol.* **1998**, *12*, 45.
- (7) Heck, S.; Kullmann, M.; Gast, A.; Ponta, H.; Rahmsdorf, H. J.; Herrlich, P.; Cato, A. C. B. *Embo Journal* **1994**, *13*, 4087.
- (8) De Bosscher, K.; Vanden Berghe, W.; Haegeman, G. *Endocr. Rev.* **2003**, *24*, 488.
- (9) Zhou, J.; Cidlowski, J. A. *Steroids* **2005**, *70*, 407.
- (10) Reichardt, H. M.; Kaestner, K. H.; Tuckermann, J.; Kretz, O.; Wessely, O.; Bock, R.; Gass, P.; Schmid, W.; Herrlich, P.; Angel, P.; Schutz, G. *Cell* **1998**, *93*, 531.
- (11) De Bosscher, K.; Vanden Berghe, W.; Haegeman, G. *J. Neuroimmunol.* **2000**, *109*, 16.
- (12) Hermoso, M.; Cidlowski, J. *IUBMB Life* **2003**, *55*, 497.
- (13) Wu, W.; Chaudhuri, S.; Brickley, D. R.; Pang, D.; Karrison, T.; Conzen, S. D. *Cancer Res.* **2004**, *64*, 1757.
- (14) Rogatsky, I.; Wang, J. C.; Derynck, M. K.; Nonaka, D. F.; Khodabakhsh, D. B.; Haqq, C. M.; Darimont, B. D.; Garabedian, M. J.; Yamamoto, K. R. *Proc. Natl. Acad. Sci USA* **2003**, *100*, 13845.
- (15) Wang, J. C.; Derynck, M. K.; Nonaka, D. F.; Khodabakhsh, D. B.; Haqq, C.; Yamamoto, K. R. *Proc. Natl. Acad. Sci USA* **2004**, *101*, 15603.
- (16) Li, X.; Wong, J.; Tsai, S. Y.; Tsai, M. J.; O'Malley, B. W. *Mol. Cell. Biol.* **2003**, *23*, 3763.

- (17) Wang, J. C.; Shah, N.; Pantoja, C.; Meijsing, S. H.; Ho, J. D.; Scanlan, T. S.; Yamamoto, K. R. *Genes Dev.* **2006**, *20*, 689.
- (18) De Bosscher, K.; Haegeman, G. *Mol. Endocrinol.* **2009**, *23*, 281.
- (19) Olenyuk, B. Z.; Zhang, G. J.; Klco, J. M.; Nickols, N. G.; Kaelin, W. G.; Dervan, P. B. *Proc. Natl. Acad. Sci USA* **2004**, *101*, 16768.
- (20) Nickols, N. G.; Jacobs, C. S.; Farkas, M. E.; Dervan, P. B. *ACS Chem. Biol.* **2007**, *2*, 561.
- (21) Nickols, N. G.; Dervan, P. B. *Proc. Natl. Acad. Sci USA* **2007**, *104*, 10418.
- (22) Dervan, P. B.; Edelson, B. S. *Curr. Opin. Struct. Biol.* **2003**, *13*, 284.
- (23) Lampson, M. A.; Kapoor, T. M. *Nat. Chem. Biol.* **2006**, *2*, 19.
- (24) Spencer, D. M.; Wandless, T. J.; Schreiber, S. L.; Crabtree, G. R. *Science* **1993**, *262*, 1019.
- (25) Stockwell, B. R. *Nat. Rev. Genet.* **2000**, *1*, 116.
- (26) Young, R. A. *Cell* **2000**, *102*, 9.
- (27) Lee, T. I.; Rinaldi, N. J.; Robert, F.; Odom, D. T.; Bar-Joseph, Z.; Gerber, G. K.; Hannett, N. M.; Harbison, C. T.; Thompson, C. M.; Simon, I.; Zeitlinger, J.; Jennings, E. G.; Murray, H. L.; Gordon, D. B.; Ren, B.; Wyrick, J. J.; Tagne, J. B.; Volkert, T. L.; Fraenkel, E.; Gifford, D. K.; Young, R. A. *Science* **2002**, *298*, 799.
- (28) Bolton, E. C.; So, A. Y.; Chaivorapol, C.; Haqq, C. M.; Li, H.; Yamamoto, K. R. *Genes Dev.* **2007**, *21*, 2005.
- (29) Ren, B.; Robert, F.; Wyrick, J. J.; Aparicio, O.; Jennings, E. G.; Simon, I.; Zeitlinger, J.; Schreiber, J.; Hannett, N.; Kanin, E.; Volkert, T. L.; Wilson, C. J.; Bell, S. P.; Young, R. A. *Science* **2000**, *290*, 2306.
- (30) Johnson, D. S.; Mortazavi, A.; Myers, R. M.; Wold, B. *Science* **2007**, *316*, 1497.
- (31) So, A. Y. L.; Chaivorapol, C.; Bolton, E. C.; Li, H.; Yamamoto, K. R. *PLOS Genetics* **2007**, *3*, 927.
- (32) Rosen, J.; Miner, J. N. *Endocr. Rev.* **2005**, *26*, 452.

- (33) Belitsky, J. M.; Nguyen, D. H.; Wurtz, N. R.; Dervan, P. B. *Bioorg. Med. Chem.* **2002**, *10*, 2767.
- (34) Nickols, N. G.; Jacobs, C. S.; Farkas, M. E.; Dervan, P. B. *Nucleic Acids Res.* **2007**, *35*, 363.
- (35) Trauger, J. W.; Dervan, P. B. *Methods Enzymol.* **2001**, *340*, 450.
- (36) Rozen, S. S., Helen J. In: Krawetz S, Misener S (eds) *Bioinformatics Methods and Protocols: Methods in Molecular Biology*. Humana Press, Totowa, NJ, pp 365-386 **2000**.
- (37) Baird, E.E.; Dervan, P.B. *J. Am. Chem. Soc.* **1996**, *118*, 6141.
- (38) (a) Iverson, B. L.; Dervan, P. B. *Nucleic Acids Res.* **1987**, *15*, 7823.
(b) Maxam, A. M.; Gilbert, W. *Methods Enzymol.* **1980**, *65*, 499.
- (39) Trauger, J. W.; Dervan, P. B. *Drug-Nucleic Acid Interactions* **2001**, *340*, 450.
- (40) Nickols, N. G. Ph. D. Thesis, California Institute of Technology, 2008. Appendix C.

Chapter 3

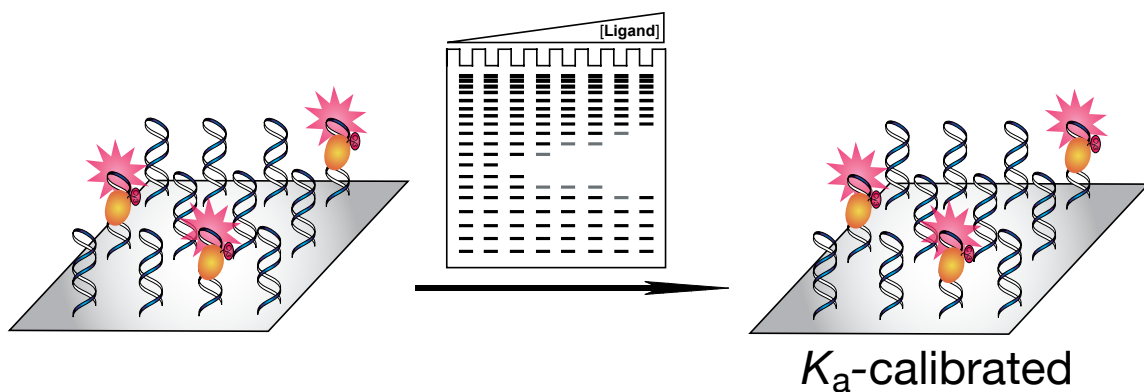
Quantitative Microarray Profiling of DNA-Binding Molecules

The text of this chapter was taken in part from a manuscript co-authored with James W. Puckett (California Institute of Technology), Josh Tietjen (University of Wisconsin, Madison), Christopher L. Warren (University of Wisconsin, Madison), Professor Aseem Z. Ansari (University of Wisconsin, Madison), and Professor Peter B. Dervan (California Institute of Technology).

(Puckett, J. W.; Muzikar, K. A.; Tietjen, J.; Warren, C. L.; Ansari, A. Z.; Dervan, P. B. "Quantitative Microarray Profiling of DNA Binding Molecules." *J. Am. Chem. Soc.* **2007**, *129*, 12310-12319.)

Abstract

A high-throughput Cognate Site Identity (CSI) microarray platform interrogating all 524,800 ten base pair-variable sites is correlated to quantitative DNase I footprinting data of DNA-binding pyrrole-imidazole polyamides. An eight-ring hairpin polyamide programmed to target the 5 base pair sequence 5'-TACGT-3' within the hypoxia response element (HRE) yielded a CSI microarray-derived sequence motif of 5'-WWACGT-3' (W = A,T). A linear β -linked polyamide programmed to target a (GAA)₃ repeat yielded a CSI microarray-derived sequence motif of 5'-AARAARWWG-3' (R = G,A). Quantitative DNase I footprinting of selected sequences from each microarray experiment enabled quantitative prediction of K_a values across the microarray intensity spectrum.



3.1 Introduction

Cell-permeable small molecules that bind specific DNA sequences and are able to interfere with protein-DNA interfaces would be useful in modulating eukaryotic gene expression. For targeting the regulatory elements of eukaryotic genes, knowledge of the preferred binding landscape of the ligand and the energetics of each site would guide gene regulation studies. Pyrrole-imidazole polyamides are a class of cell-permeable oligomers that can be programmed, based on simple aromatic amino acid pairing rules, to bind a broad repertoire of DNA sequences.¹ Knowledge of polyamide match sites has allowed us to pursue the characterization of the equilibrium association constants and hence, free energies, of hairpin polyamides for cognate DNA sites by quantitative footprint titration methods. Despite the predictive power of simple pairing rules, the sequence-dependent variability of DNA minor groove shape affords significant variability in the range of affinities for match as well as all formal single and double base pair mismatch sites.¹

Quantitative footprint titrations

Characterization of polyamide binding preferences has been studied using quantitative DNase I footprinting titrations, affording binding isotherms that enable rigorous determination of the equilibrium association constant (K_a).² The resolution of footprinting is conservatively limited to association constants of 2-fold difference or greater. Polyamide binding preferences have frequently been interrogated using DNA fragments roughly 100 base pair in size containing as many as four 6-10 base pair binding sites, which are identical with the exception of a single position that iteratively exhibits A·T, T·A, C·G, and G·C base pairs. Each binding site is interspersed with an 8 or more base pair spacer region to prevent interaction between the binding sites.³ Obtaining high quality data limits a ³²P end-labeled DNA fragment to four unique binding sites due to the resolving power of a polyacrylamide gel in a quantitative footprint titration. While DNase I footprinting has

enabled the elucidation of a binding code for hairpin polyamides, a relatively limited set of binding sites has been studied. To comprehensively interrogate all four encoded positions of an eight-ring hairpin polyamide, one would need 136 unique binding sites. In addition, interrogation of the base pairs flanking the polyamide core would necessitate 2,080 (for six base pairs total) or 32,896 (for eight base pairs total) binding sites.

CSI microarray platform

Several high-throughput platforms have been developed to characterize the binding properties of ligand-DNA interactions.⁴ Of these, two have explicitly studied the binding preferences of polyamides. The fluorescence intercalator displacement assay has interrogated polyamide binding to 512 unique five base pair sequences in a microplate format.^{4b} The more recently developed cognate site identifier (CSI) microarray platform presents all 32,896 unique eight-mers (scalable to all unique ten-mers) to fluorescently labeled polyamides, enabling an unbiased interrogation of binding preference.^{4c} By coupling DNase I footprinting with the CSI microarray data, the binding affinities (K_a values) of a given DNA-binding molecule (such as a polyamide) for any DNA sequence can be determined (**Figure 3.1**). To date, CSI microarray intensities of hairpin Cy3-polyamide conjugates have been linearly correlated to the K_a values of unlabeled polyamides.^{4c} We will examine whether this relationship between DNase I footprint titration-derived K_a values for Cy3-labeled polyamides and the corresponding microarray data remains true for additional polyamide binding architectures. Because the Cy3-polyamide conjugate may alter sequence-specificity when compared with its biologically active counterpart, the sequence-specificities of fluorophore-labeled polyamide and the biologically relevant polyamide will also be determined.

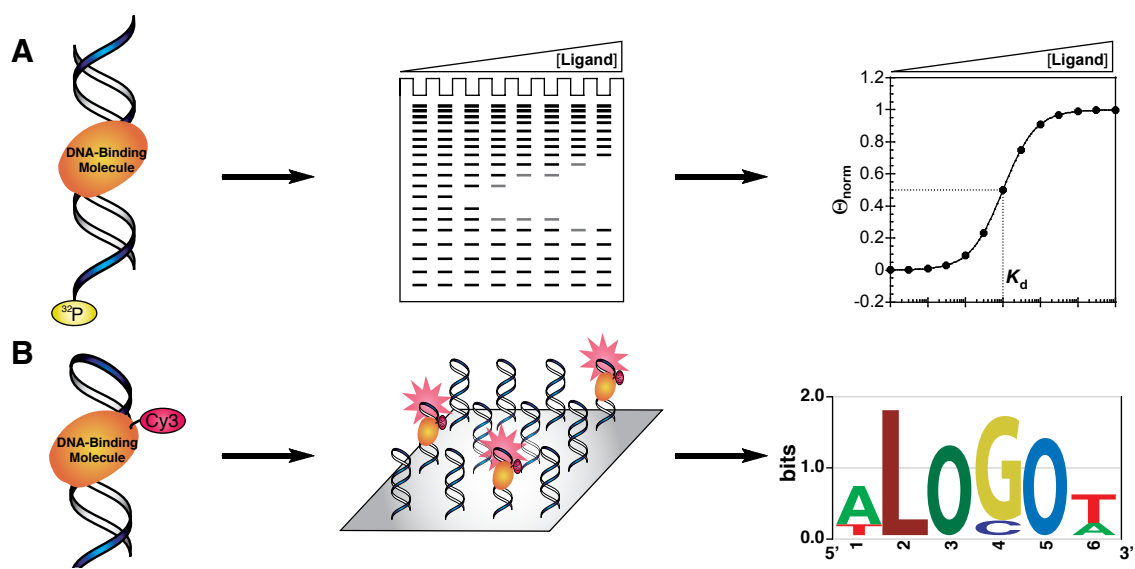


Figure 3.1 Methods for analyzing DNA binding specificity

A) Quantitative DNase I footprinting gives rise to a defined equilibrium association constant at a specified binding site for a given DNA binding molecule. B) The CSI microarray platform gives rise to relative binding preferences of an entire sequence space for the same molecule with a sequence logo as a standard summary output.

A CSI microarray harbors immense sequence specificity data; determining how to best represent this data is critical. The first reported CSI work^{4c} represented binding preferences as a sequence logo⁵ derived from several motif-finding algorithms⁶ that searched the highest Z-score bins (the ~300 highest intensities on the array), assigning equal weight to each sequence. It also examined the relative abundance of each sequence motif mutation within its respective Z-score bin.^{4c} In this paper we observe that K_a -weighting sequence motifs does not alter the sequence logo appreciably. In addition, a comprehensive single base pair mutational analysis is performed, which quantifies the specificities encoded by the polyamide at each position the polyamide interacts with DNA.

Two Cy3-labeled polyamides of biological interest⁷ are examined on the CSI microarray that displays all unique ten base pair sequences. These polyamides include a hairpin structure whose sequence specificities can be predicted from the extensive DNase I

footprinting data characterizing other pyrrole-imidazole polyamides¹ and a linear β -linked structure whose sequence specificity is less well understood.⁸ In order to correlate the CSI relative affinities (intensities) to absolute affinities (K_a values), DNase I footprinting was performed on a subset of these sequences for both the Cy3-polyamide conjugates and the related, unlabeled polyamides of known biological activity.

3.2 Results

I. Polyamide design

Two polyamide core sequences have been chosen as representative of both hairpin and linear β -linked polyamide architectures. These core recognition sequences exhibit biologically significant roles, modulating transcription in cell culture experiments.⁷ Hairpin polyamides **1** and **2** (**Figure 3.2**), were selected based on results from a project in which a polyamide-fluorescein conjugate, Ct-Py-Py-Im-(*R*)-^{H2N}- γ -Py-Im-Py-Py-(+)-FITC (**1**), displaced hypoxia inducible factor-1 α (HIF-1 α) from the hypoxia response element (HRE) of the vascular endothelial growth factor (VEGF) gene, downregulating VEGF expression 60% in cell culture experiments.^{7a,b} This eight-ring hairpin was programmed to bind the sequence 5'-WTWCGW-3' (W = A,T).^{1,3c} In particular, polyamide **1** was shown to bind the HRE sequence, 5'-TACGTG-3' on the VEGF promoter by footprint titration.^{7a,b} The Cy3 moiety was conjugated (**2**) at the same position as fluorescein for **1** to best mimic the binding properties between the two polyamides.

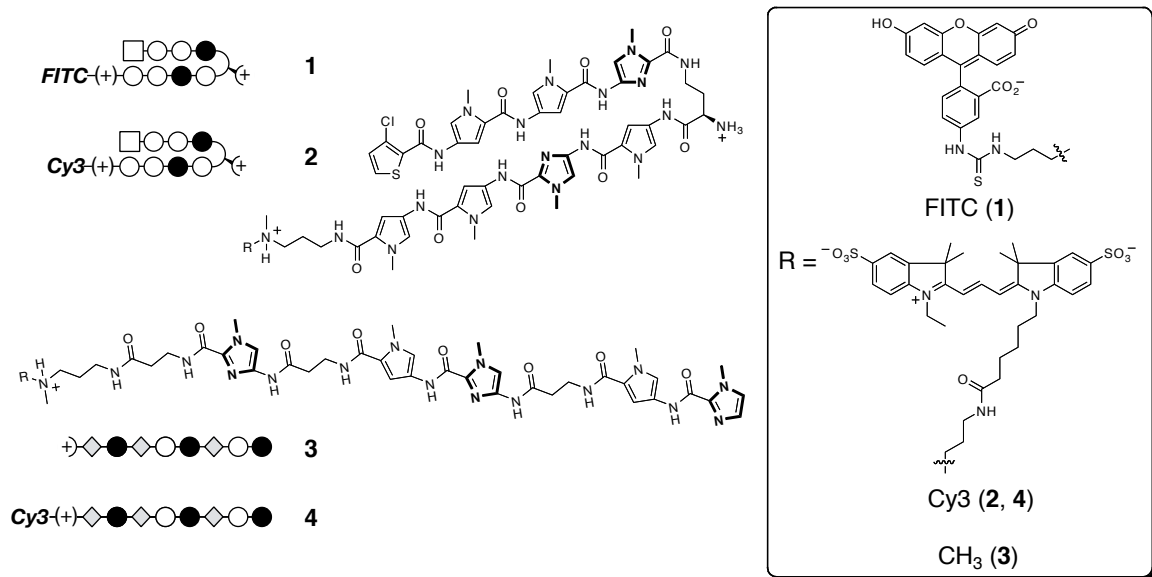


Figure 3.2 Polyamides for CSI Studies

Hairpin polyamides 1 and 2 targeted to the hypoxia response element (HRE), 5'-TACGTG-3'. Linear β-linked polyamides 3 and 4 targeted to GAA repeats in Friedreich's ataxia.

As with polyamide **1**, polyamide **3** (**Figure 3.2**) is known to bind its biologically relevant target. Polyamide **3**, Im-Py-β-Im-Py-β-Im-β-Dp, targets an intronic 5'-(GAA)_n-3' repeat hyperexpansion, enabling 2.5-fold upregulation of the *frataxin* gene, whose deficiency causes the neurodegenerative disorder Friedreich's Ataxia.^{7c} Limited knowledge about the linear β-linked class of polyamides⁸ precludes the existence of binding rules. The linear β-linked architecture has the added complexity of binding in 1:1 and 2:1 ligand/DNA stoichiometries, and we would anticipate that this class will be generally less useful due to sequence promiscuity resulting from multiple binding modes. Its 1:1 binding preferences for purine tracts, such as (GAA)_n, likely reflect shape selectivity for sequences with narrow DNA minor groove conformations.^{8c} In a 2:1 binding stoichiometry, polyamide **3** would be predicted to target 5'-WGCWGCWGCW-3'.^{8a} Remarkably, relatively few genes are affected from cell culture studies of **3** suggesting that this polyamide may be specific for 5'-AAGAAGAAG-3'.^{7c} The Cy3 fluorophore has been conjugated to the C-terminal 3,3'-diamino-N-methyldipropylamine tail (polyamide **4**).

II. CSI microarray design and results

CSI microarrays were synthesized using maskless array synthesis (MAS) technology⁹ to display all 524,800 unique ten base pair sites in quadruplicate across six microarrays. Replicates of individual hairpins occur on separate microarrays. Each hairpin on the chip consists of a self complementary palindromic sequence interrupted by a central 5'-GGA-3' sequence to facilitate hairpin formation: 5'-GCGC-N¹N²N³N⁴N⁵N⁶N⁷N⁸N⁹N¹⁰-GCGC-GGA-GCGC-N^{10'}N^{9'}N^{8'}N^{7'}N^{6'}N^{5'}N^{4'}N^{3'}N^{2'}N^{1'}-GCGC-3' (N = A,T,C,G). Previous experiments have found that 95% of the oligonucleotides on the array form duplexes.^{4c}

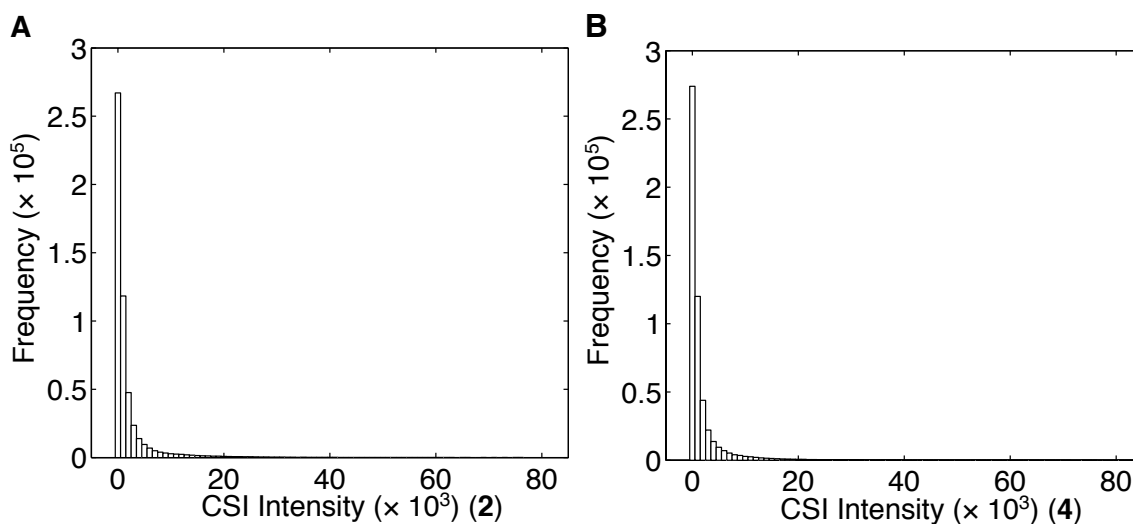


Figure 3.3 Histogram of microarray intensities

A) for polyamide **2**, B) for polyamide **4**

Polyamides **2** and **4** were slowly titrated onto the arrays and imaged at each concentration until saturation of the highest intensity binding sites was observed, 10 nM and 175 nM concentrations, respectively for **2** and **4**. After each small addition of polyamide, the arrays were washed prior to imaging. The data for each of the arrays was then normalized as previously described^{4c} to give averaged sequence intensities of the

524,800 ten base pair sites for **2** and **4**. As found with previously reported CSI arrays,^{4c} histograms of the probe intensities for **2** and **4** display a strong right-handed tail (**Figure 3.4**). The fractional standard deviations among probe replicates (standard deviation of replicates/average normalized intensity) average 0.15 ± 0.09 (polyamides **2** and **4**), for intensities exceeding 1×10^3 (**Figure 3.4**).

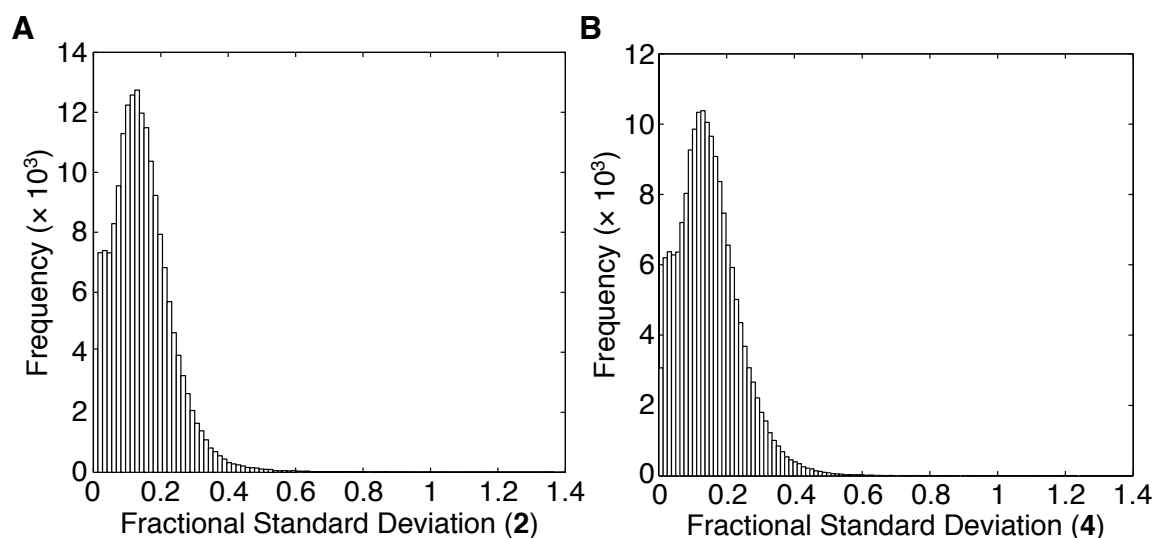


Figure 3.4 Histogram of microarray fractional standard deviations

Fractional standard deviation was calculated as standard deviation of intensities for a specific sequence/average sequence intensity for that specific sequence. A) for polyamide 2, B) for polyamide 4

III. Plasmid design

Three plasmids have been designed based on output from the CSI microarray intensities (**Figure 3.5**). Because of our interest in testing the dynamic range of the CSI assay in terms of the representative K_a values measured by a broad range of intensities, plasmids pKAM3 and pJWP17 were constructed to harbor binding sites of equal intensity spacing across a broad portion of each array's intensities, between highest and lowest intensities. The K_a values found using pKAM3 were tightly clustered across the three highest intensities,

necessitating further interrogation. Plasmid pKAM4 was designed to probe three additional intensities. A single binding site (IIIa and Ib) was held constant between pKAM3 and pKAM4 to enable interplasmid comparison of binding affinities. Because pJWP17 afforded K_a values broadly spaced across the intensity spectrum, no further study was pursued.

Since our goal is to directly compare footprinting-derived K_a values with CSI array-derived intensities, each plasmid binding site mimics the full ten base pair binding site from the array in addition to two flanking base pairs on either side of the binding site: 5'-GC-(N)₁₀-GC-3' (N = A,T,C,G). Attempts to fully replicate the 5'-GCGC-(N)₁₀-GCGC-3' binding site from the array exhibited secondary structure formation when the respective amplicons were sequenced and separated by denaturing gel electrophoresis.

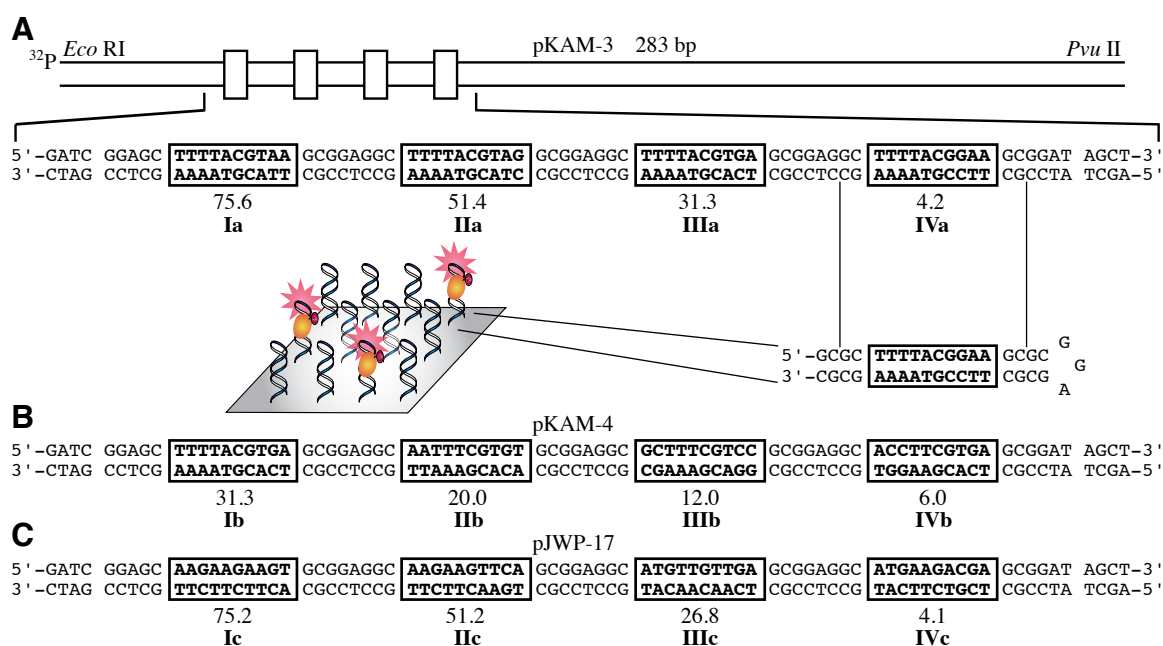




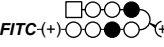
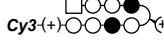
Figure 3.5 Insert sequences utilized in plasmids

Binding sites are boxed, labeled with their corresponding CSI array intensity ($\times 10^3$), and numbered. A) pKAM3 is shown, in addition to a microarray schematic demonstrating the relationship between the plasmid and a selected microarray sequence. B) pKAM4. C) pJWP17

IV. Quantitative DNase I footprint titrations: affinity and specificity determination

Hairpin polyamides **1** and **2** were incubated each for 14 h with pKAM3 or pKAM4 prior to DNase I cleavage. These two polyamides were found to bind each of seven unique ten-base pair binding sites in the same rank-order, preferentially binding 5'-TTTTACGTAA-3' with affinities of $7.5 \times 10^9 \text{ M}^{-1}$ (**1**) and $4.5 \times 10^9 \text{ M}^{-1}$ (**2**) (Table 3.1 and Figure 3.6). Replacing the fluorescein dye on polyamide **1** with Cy3 (polyamide **2**) introduced an energetic penalty that ranged from 1.5- to 10-fold, with the minimum penalty occurring at the two highest CSI intensity binding sites (Table 3.1). Polyamide **2** differentiated the highest and lowest affinity binding sites by 70-fold, slightly more than the 50-fold differentiation found for the fluorescein-labeled polyamide **1**.

Table 3.1 Quantitative DNase I footprinting-derived K_a values (M^{-1}) for **1** and **2**

pKAM-3		Ia	IIa	IIIa	IVa
Polyamide		TTTTACGTAA	TTTTACGTAG	TTTTACGTGA	TTTTACGGAA
1		$7.5 (\pm 1.8) \times 10^9$ [1]	$5.1 (\pm 0.6) \times 10^9$ [1.5]	$4.2 (\pm 0.6) \times 10^9$ [1.8]	$1.5 (\pm 0.8) \times 10^8$ [50]
2		$4.5 (\pm 1.0) \times 10^9$ [1]	$3.0 (\pm 0.6) \times 10^9$ [1.5]	$2.1 (\pm 0.3) \times 10^9$ [2.1]	$6.2 (\pm 2.0) \times 10^7$ [73]
CSI Intensity ($\times 10^3$)		75.6 (± 9.9)	51.4 (± 7.4)	31.3 (± 4.8)	4.2 (± 1.4)
pKAM-4		Ib	IIb	IIIb	IVb
Polyamide		TTTTACGTGA	AATTTTCGTGT	GCTTTTCGTCC	ACCTTCGTGA
1		$5.4 (\pm 0.9) \times 10^9$ [1.3]	$2.3 (\pm 0.1) \times 10^9$ [3.2]	$2.8 (\pm 0.2) \times 10^9$ [2.6]	$1.5 (\pm 0.2) \times 10^9$ [5]
2		$1.6 (\pm 0.2) \times 10^9$ [2.8]	$4.0 (\pm 0.9) \times 10^8$ [11]	$5.8 (\pm 0.7) \times 10^8$ [7.8]	$1.3 (\pm 0.2) \times 10^8$ [34]
CSI Intensity ($\times 10^3$)		31.3 (± 4.8)	20.0 (± 2.8)	12.0 (± 1.5)	6.0 (± 0.4)

The 10 base pair binding sites and corresponding CSI Microarray intensities are found in each column. All footprinting incubations were conducted at a minimum in triplicate at 23 °C for 14 h. Standard deviations are shown in parentheses. The bracketed numbers are $K_{a-\text{max}}/K_{a-\text{current}}$ to compare values within each polyamide series.

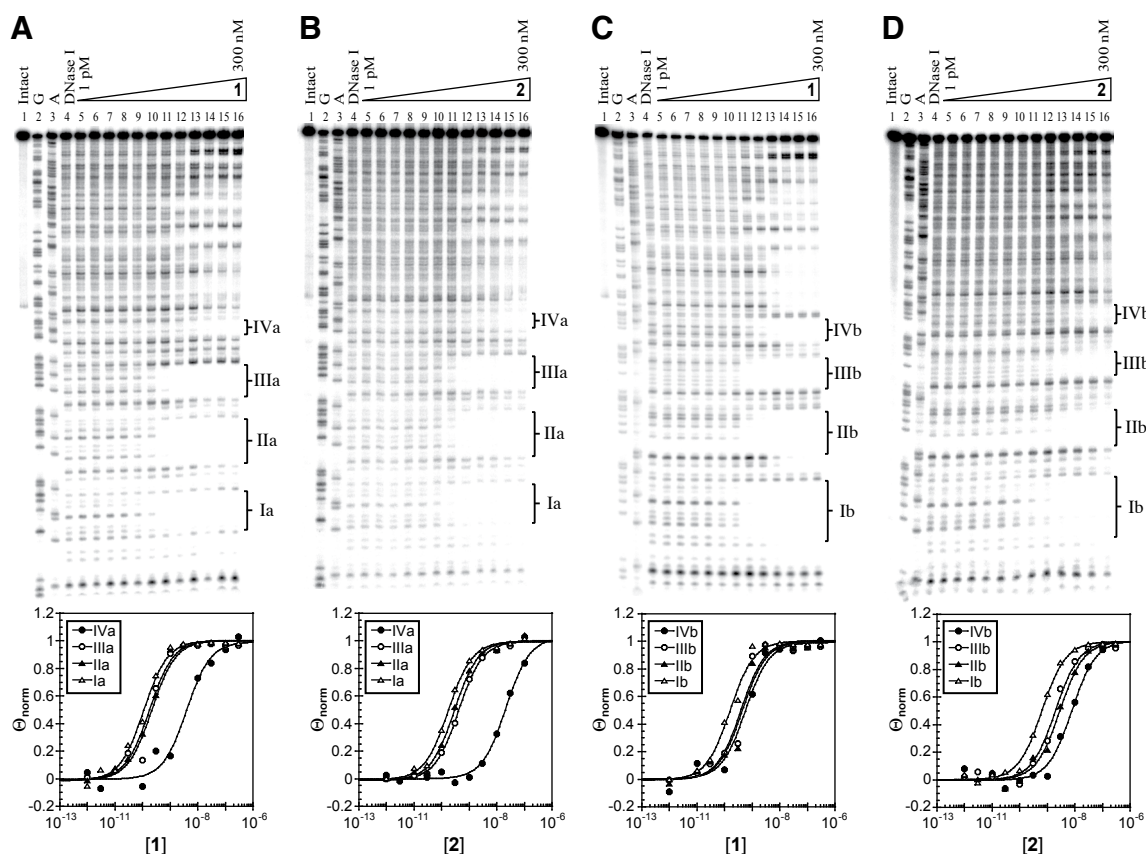


Figure 3.6 DNase I footprinting gels and corresponding isotherms of polyamides **1** and **2** on pKAM3 and pKAM4

A) Polyamide 1 on pKAM3 B) Polyamide 2 on pKAM3 C) Polyamide 1 on pKAM4 D) Polyamide 2 on pKAM4

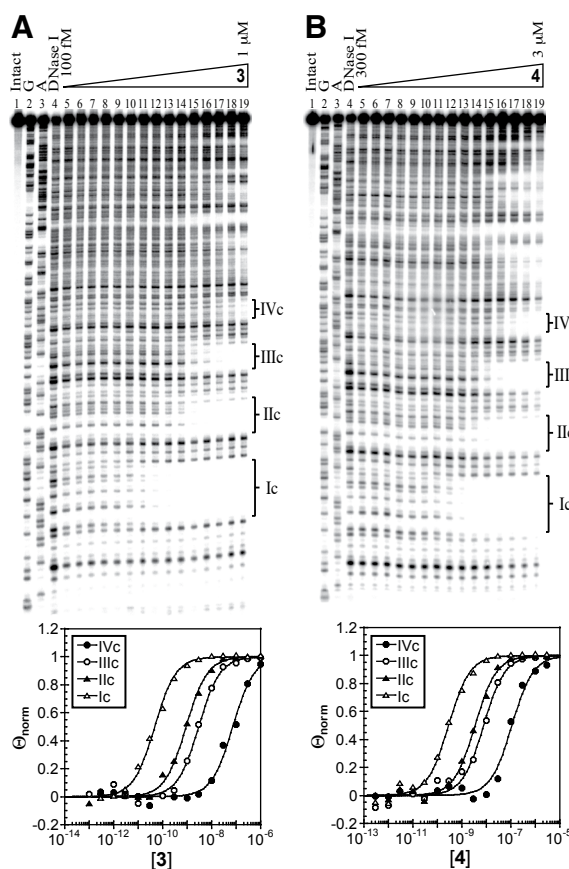
Linear β -linked polyamides **3** and **4** were each incubated for 14 h with pJWP17 prior to DNase I cleavage. They bound four unique ten base pair sites in the same rank-order, preferentially binding 5'-AAGAAGAAGT-3' (**Table 3.2** and **Figure 3.7**).

Appending the Cy3 dye to polyamide **3** either had no effect on affinity or reduced binding affinity as much as 30-fold (**Table 3.2**). Polyamide **3** bound all four binding sites over a 2400-fold range in affinity, eight times broader than for polyamide **4**.

Table 3.2 Quantitative DNase I footprinting-derived K_a values (M^{-1}) for **1** and **2**

pJWP-17		Ic	IIc	IIIc	IVc
Polyamide		AAGAAGAAGT	AAGAAGTTCA	ATGTTGTTGA	ATGAAGACGA
3	→◇●◇○●◇○●	$2.4 (\pm 0.6) \times 10^{10}$ [1]	$9.3 (\pm 2.3) \times 10^9$ [3]	$2.9 (\pm 0.7) \times 10^8$ [80]	$1.0 (\pm 0.4) \times 10^7$ [2400]
4	cy3(+)◇●◇○●◇○●	$3.3 (\pm 0.7) \times 10^9$ [1]	$2.7 (\pm 0.8) \times 10^8$ [10]	$1.1 (\pm 0.4) \times 10^8$ [30]	$1.0 (\pm 0.2) \times 10^7$ [330]
CSI Intensity ($\times 10^3$)		75.2 (± 9.2)	51.2 (± 6.2)	26.8 (± 7.3)	4.1 (± 0.4)

The 10 base pair binding sites and corresponding CSI Microarray intensities are found in each column. All footprinting incubations were conducted at a minimum in triplicate at 23 °C for 14 h. Standard deviations are shown in parentheses. The bracketed numbers are $K_{a-max}/K_{a-current}$ to compare values within each polyamide series.

**Figure 3.7** DNase I footprinting gels and corresponding isotherms of polyamides **3** and **4** on pJWP17

V. Calibrating microarrays for K_a prediction

Because DNase I footprinting enables the calculation of K_a and the direct comparison of four binding sites in a single assay, determining energetics data from CSI microarrays is crucial for understanding the global binding specificity of a polyamide. An eight-ring hairpin polyamide targeting 5'-WGWWCW-3' (W = A,T) and characterized by quantitative DNase I footprinting, Im-Py-Py-Py- γ -Im-Py-Py-Py- β -Dp,^{3c,10} has been compared to its Cy3-labeled counterpart studied on the CSI-array platform, demonstrating a linear relationship between intensity and K_a .^{4c}

Because microarray intensity at a specific microarray feature should be proportional to the fractional occupancy of DNA at that feature, the relationship between equilibrium association constant (K_a) or dissociation constant (K_d) and background-normalized microarray intensity should be:¹¹

$$\text{Intensity} = c \times \Theta = c \times \frac{K_a[\text{PA}]}{1 + K_a[\text{PA}]} = c \times \frac{[\text{PA}]}{K_d + [\text{PA}]} \quad (1)$$

In this relationship, Θ represents the fractional occupancy of DNA at a specific feature, c a scalar to reflect that microarray intensity can vary with incident laser intensity, and $[\text{PA}]$ the free polyamide concentration on the CSI array. The terms c and $[\text{PA}]$ are solved for a curve fit to Equation 1 using K_a values derived from DNase I footprint titrations and CSI microarray intensity data. Examining the limiting case where $[\text{PA}] \ll K_d$ being studied one observes a simplification to Equation 1:

$$\text{Intensity} = c \times \frac{[\text{PA}]}{K_d} = c \times [\text{PA}] \times K_a \quad (2)$$

Equation 2 represents the linear subset of the more general CSI intensity– K_a relationship described in Equation 1. Fitting the footprinting data of polyamide **2** to its corresponding microarray intensities (**Table 3.1**) using Equation 2 fits well ($R^2 = 0.94$). The linearized Equation 2 does not, however, map intensity and K_a with high correlation for polyamide **4**. Fitting the data to Equation 1 affords a better fit ($R^2 = 0.99$), indicating that $[PA]$ is not insignificant relative to the K_d of the highest intensity microarray data (**Figure 3.8**).^{12,13}

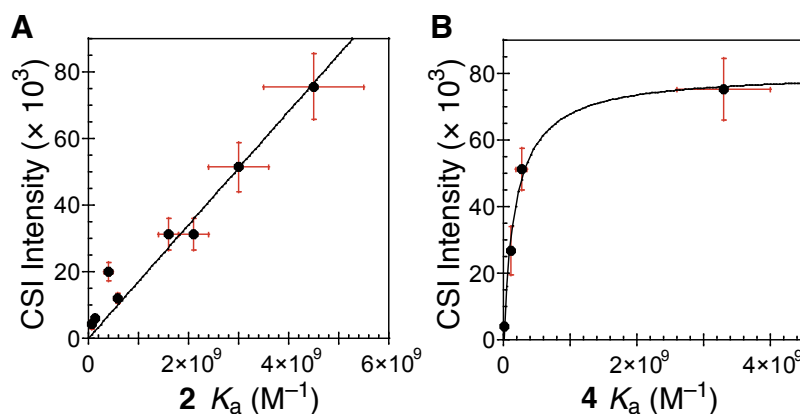


Figure 3.8 CSI array intensities correlate well with DNase I footprinting-determined K_a values

A) Polyamide **2** versus CSI array fit to Eq 2 B) Polyamide **4** versus CSI array fit to Eq 1

The K_a -calibrated microarrays can subsequently be used to interpolate K_a values from normalized sequence intensities. K_a values are derived by rearranging Equation 1 to present K_a as a function of microarray intensity:

$$K_a = \frac{\text{Intensity}}{[PA] \times (c - \text{Intensity})} \quad (3)$$

In the case where $[PA] \ll K_d$, Equation 2 rearranged to:

$$K_a = \frac{\text{Intensity}}{[PA] \times c} \quad (4)$$

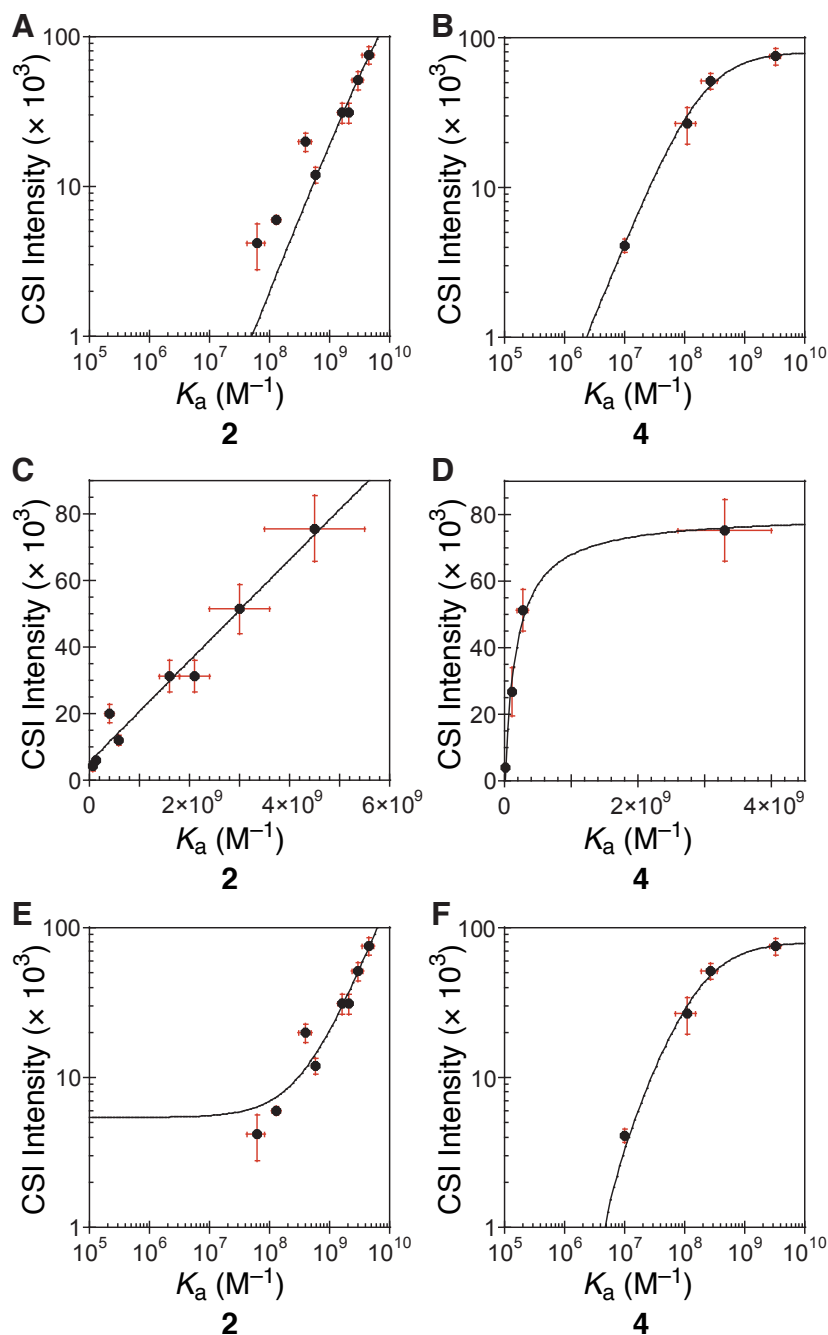


Figure 3.9 Correlation of footprinting and CSI data

A) Log-log scale plot of polyamide **2** versus CSI intensity data fit using Equation 2 B) Log-log scale plot of polyamide **4** versus CSI intensity data fit using Equation 1e. C) Linear-linear scale plot of polyamide **2** versus CSI data using Equation 2e D) Linear-linear scale plot of polyamide **4** versus CSI data using Equation 1 E) Log-log scale plot of polyamide **2** versus CSI intensity data fit using Equation 2e F) Log-log scale plot of polyamide **4** versus CSI intensity data fit using Equation 1e

VI. Correlating binding between Cy3-labeled and biologically relevant polyamides

While establishing a general K_a –intensity relationship for Cy3-labeled polyamides is a crucial first step towards global sequence interrogation of a core polyamide motif, it is equally important that the biologically relevant polyamide has sequence preferences that correlate with its Cy3-labeled counterpart. Scatter plots of polyamide **1** vs. **2** and polyamide **3** vs. **4** are best fit by a power relationship of $y=ax^n$, where (x,y) denotes the K_a values for (**1**, **2**) or (**3**, **4**) (Figure 3.10).¹⁴ The R^2 between **1** and **2** is 0.87, and that between **3** and **4** is 0.78.

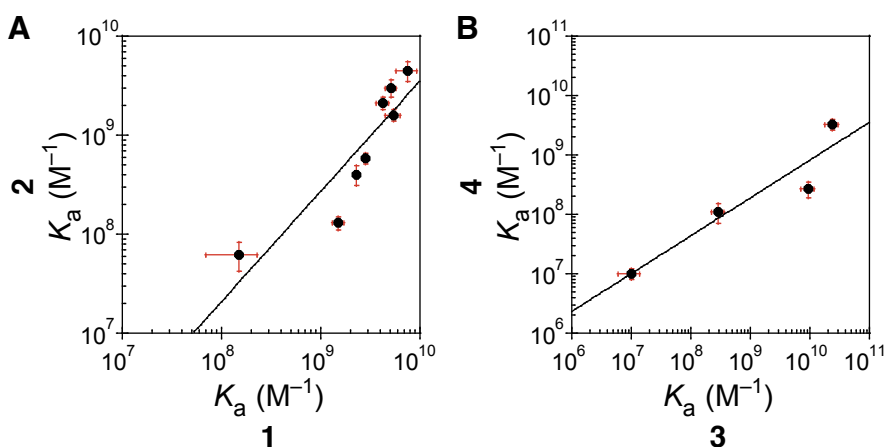


Figure 3.10 Cy3-labeled polyamides and unlabeled polyamides correlate well

A) Correlation of K_a values for polyamide **1** (fluorescein labeled) and polyamide **2** (Cy3 labeled).
 B) Correlation of K_a values for polyamide **3** (unlabeled) and polyamide **4** (Cy3 labeled)

VII. Sequence Analysis

To graphically represent the binding preferences of polyamides **2** and **4**, sequence logos have been generated (Figure 3.11 A ,B). In all cases, the motif-finding program MEME^{6a} was utilized to extract sequence motifs from the CSI binding intensities. The position-specific probability matrices output by MEME were used as inputs to enoLOGOS¹⁵

to generate a sequence logo¹⁶ The logo for polyamide **2** was created by searching the ~2500 highest sequence intensities of the CSI microarray.¹⁷ These data points span approximately a 3-fold range in K_a . The logo for polyamide **4** interrogated the 48 highest intensity sequences (a 7-fold range in K_a) of the CSI microarray.¹⁸ We examined K_a -weighted sequence logos for both polyamides **2** and **4** and found minimal differences in the resulting logos (Figure 3.11).

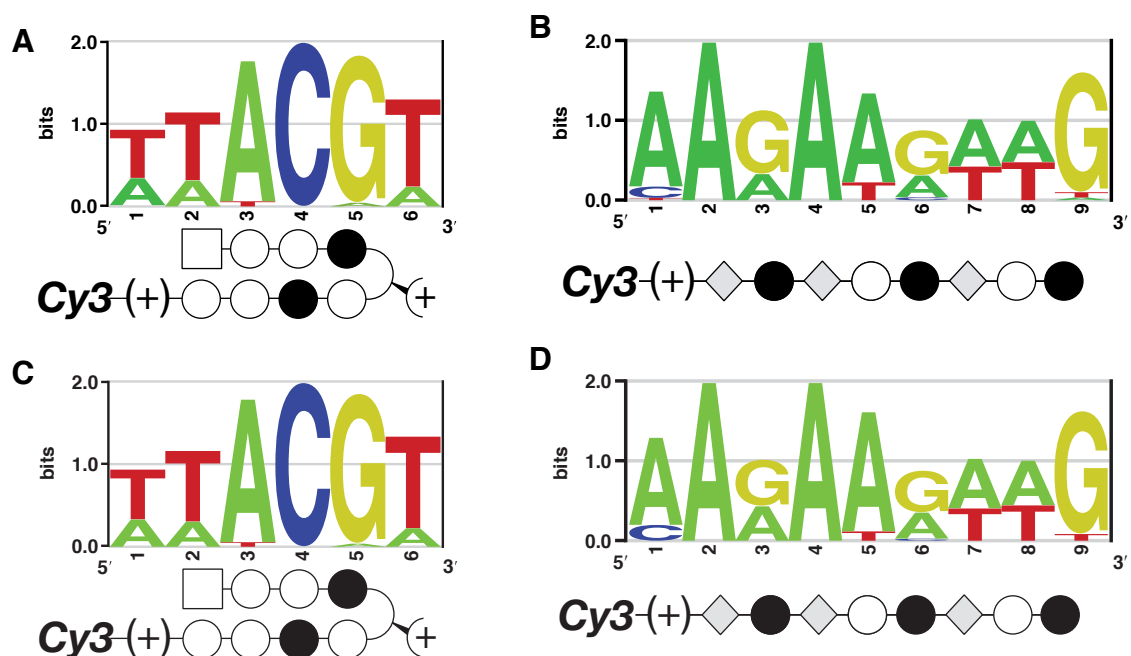


Figure 3.11 K_a -weighting components of individual sequence logos does not alter the sequence logo

A) Sequence logo for polyamide 2. B) Sequence logo for polyamide 4. C) K_a -weighted sequence logo for polyamide 2. D) K_a -weighted sequence logo for polyamide 4.

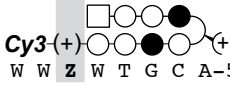
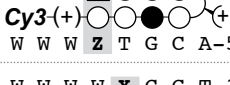

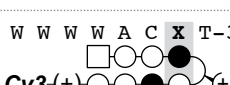
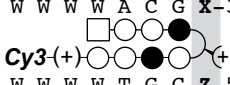
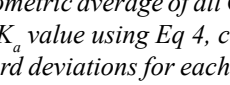
The motif for polyamide **2** has the most information at a site width of six: 5'-WWACGT-3' (Figure 3.11, W = A,T). The chlorothiophene/pyrrole pair (Ct/Py) specificity cannot be globally elucidated using polyamide **2** because of the palindromic nature of the ACGT binding site core. It is evident that the core does specify 5'-ACG-3' using Py/Py, Py/Im, and Im/Py pairings, respectively. Polyamide **3** specifies nine base pairs based on MPE footprinting data.^{8d} Polyamide **4** elicits a nine base pair motif that is best represented as

5'-AARAARWWG-3' (**Figure 3.11**, B; R = G,A and W = A,T). Previous work would suggest that Im may have no sequence preferences within linear β -linked polyamides,⁸ although this selection of nine base pair high-affinity binding sites for **4** suggests at least G·A or A·T specificity, consistent with microarray data from Friedreich's Ataxia cell culture work.^{7c}

VIII. Quantitative profiling of single base pair mismatches

While sequence logos provide a visual representation of sequence specificity, traditional studies on polyamides quantitate the specificity of a ring pairing at a selected base pair. We have examined a comprehensive single base pair mutational analysis of both polyamides **2** and **4** using K_a values interpolated from the calibrated CSI microarrays (**Table 3.3**, **Table 3.4**).¹⁹ Because the motif-finding algorithm MEME found 5'-WWACGT-3' (W = A,T) as a preferred binding sequence for polyamide **2**, we utilized this core sequence for mutational studies. Additionally, because of the 5'-ACGT-3' palindromic element of this binding site, we have isolated only binding sites containing 5'-WWWWWWACGT-3' and their mutant counterparts to preclude analyzing variants where the polyamide may be rotated 180° from the presumed orientation. To determine a K_a for 5'-WWWWWWACGT-3' (for example), the geometric mean of all microarray binding site intensities containing this motif was found. Walking from 5' to 3' on 5'-W¹W²A³C⁴G⁵T⁶-3', we observe that there is 3-fold specificity for W versus S (S = C,G) at position 1 (occupied by the linker). At position 2 (Ct/Py pair), there is 20-fold specificity for W versus S, but minimal for T·A versus A·T. The previous study of Ct/Py specificity noted only modest specificity for T·A versus A·T.^{3c} Position 3 (a Py/Py pair) confirms the previously observed W over S specificity. At position 4 (a Py/Im pair) the polyamide encodes the greatest specificity with preference for C·G versus A·T, T·A, or G·C. It is likely that this preference is at least 20-fold. At position 5, polyamide **2** appears to exhibit

Table 3.3 Microarray-derived binding affinities and specificities of all single base pair mismatch sites for polyamide **2**










Polyamide 2	X·Z	K_a (M ⁻¹)
5' -W W W W X W A C G T-3'	A·T	2.0 (1.4) × 10 ⁹
	T·A	2.5 (1.2) × 10 ⁹
3' -W W W W Z T G C A-5'	C·G	6.9 (1.4) × 10 ⁸
3' -W W W W Z T G C A-5'	G·C	6.8 (1.6) × 10 ⁸
5' -W W W W W X A C G T-3'	A·T	1.8 (1.3) × 10 ⁹
	T·A	2.7 (1.2) × 10 ⁹
3' -W W W W W Z T G C A-5'	C·G	1.0 (2.0) × 10 ⁸
3' -W W W W W Z T G C A-5'	G·C	1.3 (2.2) × 10 ⁸
5' -W W W W W W X C G T-3'	A·T	2.2 (1.3) × 10 ⁹
	T·A	1.1 (1.6) × 10 ⁹
3' -W W W W W W Z G C A-5'	C·G ^a	≤ 10 ⁸
3' -W W W W W W Z G C A-5'	G·C	1.3 (2.5) × 10 ⁸
5' -W W W W W W A X G T-3'	A·T ^a	≤ 10 ⁸
	T·A ^a	≤ 10 ⁸
3' -W W W W W W T Z C A-5'	C·G	2.2 (1.3) × 10 ⁹
3' -W W W W W W T Z C A-5'	G·C ^a	≤ 10 ⁸
5' -W W W W W W A C X T-3'	A·T	1.2 (1.4) × 10 ⁹
	T·A	2.9 (1.8) × 10 ⁸
3' -W W W W W W T G Z A-5'	C·G	2.4 (1.8) × 10 ⁸
3' -W W W W W W T G Z A-5'	G·C	2.2 (1.3) × 10 ⁹
5' -W W W W W W A C G X -3'	A·T	1.3 (1.4) × 10 ⁹
	T·A	2.2 (1.3) × 10 ⁹
3' -W W W W W W T G C Z -5'	C·G ^a	≤ 10 ⁸
3' -W W W W W W T G C Z -5'	G·C ^a	≤ 10 ⁸

All K_a values are derived from the geometric average of all CSI binding site intensities on the array containing a specified sequence, converted to a K_a value using Eq 4, corrected to include an error term ϵ . The values in parentheses are the geometric standard deviations for each K_a value. X·Z entries marked with a superscripted “^a” contain averaged intensities below ϵ . For these entries, an upper bound on the K_a is estimated based on the log-log plot of K_a versus intensity found in **Figure 3.9**.

less specificity than would be predicted for an Im/Py ring pair, binding almost as well to A·T as to G·C.¹ The polyamide “turn unit,” position 6, confirms a strong preference for W over S.¹ Through this quantitative study, we observe four strongly encoded binding positions, italicized in 5'-*WWCGW*-3'. The discrepancy between the observed sequence logo, as found by MEME, and the suggested specificity by a single base pair mutation study likely stems from (i) the examination of all sequences in the single base pair mutation as compared to only a subset for the sequence logo, (ii) the assumption by the logo of independence of base pair-polyamide interaction at each position, and (iii) the examination in the single base pair mutation of the average K_a of a group of sequences containing a

specified motif.

Table 3.4 Microarray-derived binding affinities and specificities of all single base pair mismatch sites for polyamide 4

Polyamide 4	X·Z	K_a (M ⁻¹)
5' - X A R A A R W W G-3' Cy3 (+)-  3' - Z T Y T T Y W W C-5'	A·T T·A C·G G·C	$1.6 (2.2) \times 10^8$ $8.7 (2.0) \times 10^7$ $4.3 (1.8) \times 10^7$ $4.7 (1.8) \times 10^7$
5' -A X R A A R W W G-3' Cy3 (+)-  3' -T Z Y T T Y W W C-5'	A·T T·A C·G G·C	$1.6 (2.2) \times 10^8$ $7.7 (2.2) \times 10^7$ $2.1 (1.8) \times 10^7$ $3.2 (1.9) \times 10^7$
5' -A A X A A R W W G-3' Cy3 (+)-  3' -T T Z T T Y W W C-5'	A·T T·A C·G G·C	$1.3 (2.0) \times 10^8$ $3.5 (1.7) \times 10^7$ $8.4 (1.6) \times 10^7$ $2.0 (2.2) \times 10^8$
5' -A A R X A R W W G-3' Cy3 (+)-  3' -T T Y Z T Y W W C-5'	A·T T·A C·G G·C	$1.6 (2.2) \times 10^8$ $4.3 (1.8) \times 10^7$ $9.9 (1.9) \times 10^6$ $7.5 (2.6) \times 10^6$
5' -A A R A X R W W G-3' Cy3 (+)-  3' -T T Y T Z Y W W C-5'	A·T T·A C·G G·C	$1.6 (2.2) \times 10^8$ $8.3 (2.2) \times 10^7$ $9.1 (2.1) \times 10^6$ $1.1 (2.0) \times 10^7$
5' -A A R A A X W W G-3' Cy3 (+)-  3' -T T Y T T Z W W C-5'	A·T T·A C·G G·C	$1.5 (2.1) \times 10^8$ $5.5 (1.7) \times 10^7$ $8.5 (1.7) \times 10^7$ $1.7 (2.3) \times 10^8$
5' -A A R A A R X W G-3' Cy3 (+)-  3' -T T Y T T Y Z W C-5'	A·T T·A C·G G·C	$1.7 (2.2) \times 10^8$ $1.5 (2.2) \times 10^8$ $2.2 (2.2) \times 10^7$ $2.0 (2.4) \times 10^7$
5' -A A R A A R W X G-3' Cy3 (+)-  3' -T T Y T T Y W Z C-5'	A·T T·A C·G G·C	$1.6 (2.5) \times 10^8$ $1.5 (2.0) \times 10^8$ $3.5 (2.0) \times 10^7$ $3.9 (2.0) \times 10^7$
5' -A A R A A R W W X -3' Cy3 (+)-  3' -T T Y T T Y W W Z -5'	A·T T·A C·G G·C	$1.2 (1.9) \times 10^8$ $1.1 (2.0) \times 10^8$ $1.1 (1.7) \times 10^8$ $1.6 (2.2) \times 10^8$

All K_a values are derived from the geometric average of all CSI binding site intensities on the array containing a specified sequence, converted to a K_a value using Equation 3, corrected to include an error term ϵ . The values in parentheses are the geometric standard deviations for each K_a value.

3.3 Discussion

In conjunction with the sequence logo for polyamide **2**, the CSI array analysis validates the sequence specificity programmed by the aromatic amino acid ring pairs. The extensive DNase I footprinting data on eight-ring and six-ring hairpin polyamides, while limited on the scale of a CSI microarray, enabled the creation of pairing rules that are remarkably general.¹ It is evident from the microarray that Im/Py and Py/Im ring pairs offer the greatest specificity for a single base pair, while Py/Py, Ct/Py, and the “turn unit” afford general W specificity. While the Ct/Py ring pair conferred minimal specificity for T·A versus A·T, its W specificity is likely an improvement over the use of a Py/Py ring pair, which at the N-terminus of an eight-ring hairpin polyamide exhibits specificity for A·T, T·A, and G·C versus C·G.^{3c} The sequence specificity of **2** correlates remarkably well with the 5'-ACGT-3' specificity of echinomycin,²⁰ also known to affect VEGF expression in cell culture.²¹ The examination of polyamide **4** marks the most comprehensive sequence specificity study of a linear b-linked polyamide since the original examination of the binding specificity for Im- β -Im-Py- β -Im- β -Im-Py- β -Dp.^{8a,b} In the 5'-A¹A²R³A⁴A⁵R⁶W⁷W⁸G⁹-3' sequence (R = G,A; W = A,T), positions **4**, **5**, and **7**, each containing either a Py or a β , exhibit the greatest specificity for W over S (S = C,G). Intriguingly, the β at position 4 prefers A·T over T·A, an unexpected specificity. The sequence logo for polyamide **4** indicates that Im has modest preference for G·C or A·T over other base pairings – in this mutational study, however, imidazole is generally degenerate. The wide range of K_a values comprising each motif (high geometric standard deviation) make the statistical significance of any specificities under 4 relatively small. In general, the geometric standard deviations for polyamide **4** were higher than those for polyamide **2**, when including only those table entries for polyamide **2** in which each K_a value was composed of all instances of the motif. One potential source of the increased standard deviation in binding affinities is the single variable base flanking the nine base pair binding site for polyamide **4**. Because the minor

groove width is a potentially important contributor to binding affinity and specificity for the linear β -linked class of polyamides,^{8c} a single variable, flanking base is unlikely to enable comprehensive interrogation of the global set of sequence-dependent DNA microstructures. As with polyamide **2**, the discrepancies observed between the sequence logo of polyamide **4** and the comprehensive single base pair mutational analysis likely stem from similar causes.

With the sequence logo (approximated as 5'-AARAARWWG-3') as a snapshot of the highest affinity binding sites for polyamide **4** ($K_a \approx 5 \times 10^8$ to $3.3 \times 10^9 \text{ M}^{-1}$) and the footprint titration binding isotherms for determining DNA binding mode, we confirm a preference for the 1:1 binding stoichiometry. Previous data characterizing the linear β -linked polyamide Im- β -Im-Py- β -Im- β -Im-Py- β -Dp demonstrated a 30-fold energetic preference for the 1:1 versus 2:1 binding stoichiometry, presumably due to the increased entropic cost of the 2:1 binding mode.^{8a} It is remarkable that polyamide **3** exhibited specificity for upregulation of the *frataxin* gene in cell culture,^{7c} since the sequence preference for **4** was not overwhelmingly 5'-AAGAAGAAG-3'. Two possible explanations for this observation are (i) that multiple binding events in the genome have marginal effects on transcription and that the specificity is amplified by the GAA repeat expansion in Friedreich's Ataxia, or (ii) that many of the sequences described by 5'-AARAARWWG-3' exist in higher order chromosomal structures that cannot be targeted by polyamide **3**.

Suggestions for microarray usage

In the case where the free ligand concentration is small relative to the K_d for each binding site on the CSI microarray, a linear K_a – intensity relationship is observed. The binding profiles examined for polyamide **2** and for previously studied molecules are examples of linear K_a – intensity relationships. For the highest intensity sites also studied by DNase I footprinting (**Figure 3.8a**), the CSI microarray experiment contains greater

resolving power and can differentiate K_a values that are indistinguishable by quantitative DNase I footprint titrations. In this example, as CSI intensity data approach ϵ , small changes in intensity yield large changes in predicted K_a . Because the characterization of DNA-binding ligands is most concerned with defining a perfect match site, this limitation is minor. CSI data for polyamide **2** conservatively enables distinguishing a 50-fold range of K_a values, thus encompassing the majority of single base pair mismatch specificities.

In the case where the free ligand concentration is comparable to the K_d , a non-linear K_a –intensity relationship is observed. The binding profile for polyamide **4** marks an example of a CSI microarray studied compound that occurs outside the linear range of Equation 1. In this case, closely clustered high-intensity data points can span a broad range of K_a values (**Figure 3.8b**). The error inherent to the CSI microarray analysis is thus amplified when K_a values in this high-CSI-intensity region are interpolated.

Because of the gradual polyamide titration onto the array, it should be possible to capture snapshots of both polyamide saturation within the linear K_a –intensity region for the highest affinity binding sites on the microarray and binding site saturation enabling lower intensity data points to fall within the higher precision linear K_a –intensity region. Such titration may enable high-precision K_a data to be extracted from all intensities of the microarray.

The sequence logos presented in this paper represent a snapshot of a binding profile for the highest affinity binding sites by a dye-labeled ligand. The polyamide core dictates the majority of the binding specificity revealed by CSI microarray analysis – the presence of a Cy3 label may reduce affinity to a binding site relative to its unlabeled counterpart but does not alter the rank-order of binding preferences. Complementing the graphical image of a sequence logo, the comprehensive single base pair mutational analysis afforded by

the extensive microarray data quantitates one's understanding of the polyamide sequence preferences.

Conclusion

Correlating the sequence preference landscape present on the CSI microarray to quantitative footprinting enables energetic studies using global binding information. This capacity marks a significant forward step for the field of small molecule·DNA recognition and enables the comprehensive interrogation of DNA binding small molecules to be better understood. The elucidation of 5'-WWACGT-3' as the binding site for **2** confirmed the previously established pairing code for hairpin polyamides, and the determination of 5'-AARAARWWG-3' for **4** helps explain the specificity it exhibited in cell culture. The correlation between a Cy3-labeled polyamide and an unlabeled polyamide of biological interest means that these motifs well approximate the binding profiles for **1** and **3**, respectively. DNase I footprinting-calibrated CSI microarrays have been shown to be an effective technique for determining the binding affinities of DNA-binding ligands for a vastly expanded repertoire of DNA sequences, and we envision them to be a critical tool for reliably determining sequence specificity for other ligands in the future.

3.4 Materials and Methods

Materials

Boc- β -Ala-Pam resin (0.81 mmol/g), Boc₂O and the chiral amino acid Fmoc-D-Dab-Boc-OH were purchased from Peptides International. Kaiser oxime resin (0.56 mmol/g) was purchased from Novabiochem. Trifluoroacetic acid (TFA) was purchased from Halocarbon. All solvents were purchased from Aldrich or EMD Biosciences. rac-Dithiothreitol (DTT) was purchased from ICN. Cy3-NHS ester was purchased from Invitrogen. RNase-free DEPC water was purchased from US Biochemicals. Water (18.2 M Ω) was purified using a Millipore water purification system.

The pH of buffers was adjusted using a Beckman 340 pH/temp meter. All buffers were sterilized by filtration through either a Nalgene 0.2 μ m cellulose nitrate filtration device or a 0.2 μ m Whatman cellulose acetate disposable syringe filter. DNA oligonucleotides were ordered PAGE-purified from Integrated DNA Technologies. [γ -³²P] adenosine 5'-triphosphate (≥ 7000 Ci/mmol) was obtained from MP Biomedicals. Sonicated, phenol-extracted calf thymus DNA was from Amersham, and all enzymes and molecular biology grade glycogen (20 mg/mL) were purchased from Roche.

Methods

UV spectra were recorded in water using an Agilent 8453 UV-Vis spectrophotometer for polyamide **3**. To solvate polyamides **1**, **2**, and **4** in water at sufficiently high concentrations for DNase I footprinting, 5 mL of DMSO was added to a 20 nmol aliquot of polyamide. The solution was then diluted using RNase-free DEPC water. UV spectra of **1**, **2**, and **4** were blanked against solutions containing appropriate amounts of DMSO. The concentrations of polyamides **1** and **2** were determined using a local λ_{max} of 313 nm, $\epsilon = 69,500 \text{ L}\cdot\text{mol}^{-1}\cdot\text{cm}^{-1}$. The concentrations of polyamides **3** and **4** were determined

using a local λ_{max} of 288 nm, $\epsilon = 43,125 \text{ L}\cdot\text{mol}^{-1}\cdot\text{cm}^{-1}$. LASER desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) was performed using an Applied Biosystems Voyager DE Pro spectrometer. Analytical and preparative high-pressure liquid chromatography (HPLC) were performed with a Beckman Gold system equipped with a diode array (analytical) or single-wavelength (preparative) detector as previously described.^{4c,7}

i. Polyamide synthesis

Pyrole and imidazole monomer units (Boc-Py-OBt [(1,2,3-Benzotriazol-1-yl 4-[(tert-Butoxycarbonyl)amino]-1-methylpyrrole-2-carboxylate) and Boc-Im-OH (4-[(tert-Butoxycarbonyl)amino]-1-methylimidazole-2-carboxylic acid) respectively) are synthesized according to established protocols²⁵ and maintained as general group stock. Fmoc-D(Dab)-Boc-OH turn monomer unit and Boc₂O were purchased from Peptides International.

Polyamides **1** and **2** were synthesized on Kaiser oxime resin²² and **3** and **4** were synthesized on Boc- β -Ala-PAM resin,²³ using previously described methods.^{4c,7} In a typical synthesis, 4 molar equivalents (relative to resin loading for the lot of oxime resin used) of monomer unit were activated with 4 equivalents of PyBOP (NovaBiochem) and 4 equivalents of diisopropylethylamine (DIEA) in a solution of dimethylformamide (DMF) (solution was approximately 0.3-0.5 M in monomer unit). In a glass peptide synthesis vessel, the resin was shaken with this solution for 2-4 hours at room temperature until the reaction was complete, as assessed by analytical HPLC of a cleaved sample of resin. Resin was deprotected by shaking in a solution of 20% trifluoroacetic acid (TFA) in methylene chloride for 20 minutes at room temperature. After completion of synthesis of the oligomer the Fmoc protecting group was removed by 10 minute incubation with 20% piperidine in

DMF at room temperature. The free amine was subsequently protected by incubation of the resin for 1 hour in a solution of Boc₂O in DMF. The polyamide was subsequently cleaved from resin by incubation at 55 °C with 3,3-diamino-N-methyl-dipropylamine overnight (8-14 hours). The cleavage product was purified by reverse-phase HPLC, frozen in LN₂ and lyophilized to dryness.

To conjugate polyamides to FITC or Cy3 fluorophore, a lyophilized aliquot of polyamide was dissolved in DMF to a final concentration ~0.03M, DIEA (100 eq) and 1mg FITC isomer 1 isocyanate (polyamide **1**) or Cy3-NHS monoester (polyamides **2** and **4**) were added and the reaction was allowed to shake in darkness for 1 hour at 37 °C. Following conjugation the polyamide was precipitated from solution by addition of cold diethyl ether and the precipitate was pelleted by centrifugation, the supernatant was removed and the pellet was allowed to air-dry. The Boc protecting group was removed by the addition of 100μL of neat TFA to the pellet. The deprotection proceeded for 10 min at room temperature before being diluted with 20% acetonitrile/water (0.1% TFA) followed by purification by reverse-phase HPLC. Purities and identities of the polyamides were assessed by HPLC, UV-visible spectroscopy, and MALDI-TOF MS.

1: (MALDI-TOF) [M+H]⁺ calcd for C₇₇H₈₀ClN₂₂O₁₄S₂⁺ 1635.5, observed 1635.9

2: (MALDI-TOF) [M+H]⁺ calcd for C₈₇H₁₀₅ClN₂₃O₁₆S₃⁺ 1858.7, observed 1858.7

3: (MALDI-TOF) [M+H]⁺ calcd for C₄₁H₅₆N₁₇O₈⁺ 914.4, observed 914.4

4: (MALDI-TOF) [M+H]⁺ calcd for C₇₄H₇₄N₂₀O₁₅S₂⁺ 1569.7, observed 1569.6

ii. Plasmid preparation

Plasmids were constructed by ligating the following hybridized inserts into the BamHI / HindIII polycloning site in pUC19:

pKAM3:

5'-GATCGGAGCTTTTACGTAAGCGGAGGCTTTTACGTAGGCGGAGGCTTTTAC-
GTGAGCGGAGGCTTTTACGGAAGCGGAT-3'

5'-AGCTATCCGCTTCCGTAAAAGCCTCCGCTCACGTAAAAGCCTCCGCCTACG-
TAAAAGCCTCCGCTTACGTAAAAGCTCC-3'

pKAM4:

5'-GATCGGAGCTTTTACGTGAGCGGAGGCAATTTTCGTGTGCGGAGGCGCTTTC-
GTCCGCGGAGGCACCTTCGTGAGCGGAT-3'

5'-AGCTATCCGCTCACGAAGGTGCCTCCGCGGACGAAAGCGCCTCCGCACAC-
GAAATTGCCTCCGCTCACGTAAAAGCTCC-3'

pJWP17:

5'-GATCGGAGCAAGAAGAAGTGCGGAGGCAAGAAGTTCAGCGGAGGCAT-
GTTGTTGAGCGGAGGCATGAAGACGAGCGGAT-3'

5'-AGCTATCCGCTCGTCTTCATGCCTCCGCTCAACAACATGCCTCCGCT-
GAACTTCTTGCCTCCGCACTTCTTCTTGCTCC-3.'

The ligated plasmid was then transformed into JM109 subcompetent cells (Promega) by standard methods (30 minute incubation on ice followed by 45 second heat shock (42 °C) followed by 1 hour incubation at 37 °C). Colonies were selected for α -complementation on agar plates containing 50 mg/L ampicillin, 120 mg/L IPTG, and 40 mg/L X-gal after overnight growth at 37 °C. Cells were harvested after 16 h growth at 37 °C in LB medium containing 50 mg/L ampicillin. Plasmids were then purified by mini-prep kits. The presence of the desired inserts was determined by capillary electrophoresis dideoxy sequencing methods.

iii. Preparation of 5'-labeled DNA for DNase I footprinting

Two primer oligonucleotides, 5'-AATTCGAGCTCGGTACCCGGG-3' (forward) and 5'-CTGGCACGACAGGTTTCCCGA-3' (reverse) were constructed for PCR amplification. The forward primer was radiolabeled using [γ -³²P]-dATP and polynucleotide kinase, followed by purification using ProbeQuant G-50 spin columns. The desired DNA region was amplified as previously described.² The labeled fragment was loaded onto a 7% nondenaturing preparatory polyacrylamide gel (5% cross-link), and the desired 283 (pKAM3, pKAM4, pJWP17) base-pair band was visualized by autoradiography and isolated. Chemical sequencing reactions were performed according to published protocols.²⁶

iv. Quantitative DNase I footprint titrations

All reactions were carried out in a volume of 400 μ L according to published protocols. Polyamides were equilibrated with the radiolabeled DNA for 14 h prior to DNase I cleavage at 23 °C. Quantitation by storage phosphor autoradiography and determination of equilibrium association constants were as previously described.²

v. Microarray procedures

Microarrays were synthesized by using a Maskless Array Synthesizer (NimbleGen Systems, Madison, WI). Homopolymer (T_{10}) linkers were covalently attached to monohydroxysilane glass slides. Oligonucleotides were then synthesized on the homopolymers to create a high-density oligonucleotide microarray. The array surface was derivatized such that the density of oligonucleotides was sufficiently low within the same feature so that no one oligonucleotide would hybridize with its neighbors. Four copies of each hairpin containing a unique ten base pair site (5'-GCGC-N¹N²N³N⁴N⁵N⁶N⁷N⁸N⁹N¹⁰-GCGC-GGA-GCGC-N^{10'}N^{9'}N^{8'}N^{7'}N^{6'}N^{5'}N^{4'}N^{3'}N^{2'}N^{1'}-GCGC-3') required a total of 2,099,200 features, divided among six microarrays.

vi. Binding assay

Microarray slides were immersed in 1x PBS and placed in a 90 °C water bath for 30 min to induce hairpin formation of the oligonucleotides. Slides were then transferred to a tube of nonstringent wash buffer (saline/sodium phosphate/EDTA buffer, pH 7.5/0.01% Tween 20) and scanned to check for low background (<200 intensity). Microarrays were scanned by using an Axon 4000B, and the image files were extracted with GENEPIX PRO Version 3.0 (Axon Instruments, Foster City, CA).

vii. Polyamide binding

Microarrays prepared as above were placed in the microarray hybridization chamber and washed twice with nonstringent wash buffer. Polyamide was diluted to 10 nM (for **2**) or 175nM (for **4**) in Hyb buffer (100 mM Mes/1 M NaCl/20 mM EDTA, pH 7.5/0.01% Tween 20). Polyamide was then added to the hybridization chamber and incubated at room

temperature for 1 h. Finally, the microarrays were washed twice with nonstringent wash buffer and scanned.

viii. Data processing

For each replicate, global mean normalization was used to ensure the mean intensity of each microarray was the same. Local mean normalization²⁴ was then used to ensure that the intensity was evenly distributed throughout each sector of the microarray surface. Outliers between replicate features were detected by using the Q test at 90% confidence and filtered out. The replicates were then quantile-normalized²⁵ to account for any possible nonlinearity between arrays. Duplicate features were then averaged together. The median of the averaged features was subtracted to account for background.

3.5 Acknowledgements

This work was supported by the National Institutes of Health (GM27681 to P.B.D. and A.Z.A). We thank the Beckman Institute Sequence Analysis Facility for DNA sequencing.

3.6 References

- (1) Im / Py targets G·C; Py / Py targets A·T and T·A; and Ct / Py targets T·A. (a) Dervan, P. B.; Edelson, B. S. *Curr. Opin. Struct. Biol.* **2003**, *13*, 284. (b) Hsu, C. F.; Phillips, J. W.; Trauger, J. W.; Farkas, M. E.; Belitsky, J. M.; Heckel, A.; Olenyuk, B. Z.; Puckett, J. W.; Wang, C. C. C.; Dervan, P. B. *Tetrahedron* **2007**, *63*, 6146.
- (2) (a) Trauger, J. W.; Dervan, P. B. *Methods Enzymol.* **2001**, *340*, 450.
- (3) (a) Doss, R. M.; Marques, M. A.; Foister, S.; Chenoweth, D. M.; Dervan, P. B. *J. Am. Chem. Soc.* **2006**, *128*, 9074. (b) Marques, M. A.; Doss, R. M.; Foister, S.; Dervan, P. B. *J. Am. Chem. Soc.* **2004**, *126*, 10339. (c) Foister, S.; Marques, M. A.; Doss, R. M.; Dervan, P. B. *Bioorg. Med. Chem.* **2003**, *11*, 4333.
- (4) (a) For a review on Protein Binding Microarrays (PBMs), see: Bulyk, M. L. *Methods Enzymol.* **2006**, *410*, 279. (b) For a review on fluorescence intercalator displacement (FID) assays, see: Tse, W. C.; Boger, D. L. *Acc. Chem. Res.* **2004**, *37*, 61. (c) For the initial report of cognate site identifier (CSI) microarrays, see: Warren, C. L.; Kratochvil, N. C. S.; Hauschild, K. E.; Foister, S.; Brezinski, M. L.; Dervan, P. B.; Phillips, G. N.; Ansari, A. Z. *Proc. Natl. Acad. Sci USA* **2006**, *103*, 867.
- (5) Schneider, T. D.; Stephens, R. M. *Nucleic Acids Res.* **1990**, *18*, 6097.
- (6) (a) Bailey, T. L.; Elkan, C. In *Proceedings of the Second International Conference on Intelligent Systems for Molecular Biology*; Altman, R., Brutlag, D., Karp, P., Lathrop, R., Searls, D., Eds.; AAAI Press: Menlo Park, 1994, pp 28. (b) Liu, X. S.; Brutlag, D. L.; Liu, J. S. *Nat. Biotechnol.* **2002**, *20*, 835. (c) Hughes, J. D.; Estep, P. W.; Tavazoie, S.; Church, G. M. *J. Mol. Biol.* **2000**, *296*, 1205.
- (7) (a) Olenyuk, B. Z.; Zhang, G. J.; Klco, J. M.; Nickols, N. G.; Kaelin, W. G.; Dervan, P. B. *Proc. Natl. Acad. Sci USA* **2004**, *101*, 16768. (b) Nickols, N. G.; Jacobs, C. S.; Farkas, M. E.; Dervan, P. B. *Nucleic Acids Res.* **2007**, *35*, 363. (c) Burnett, R.; Melander, C.; Puckett, J. W.; Son, L. S.; Wells, R. D.; Dervan, P. B.; Gottesfeld, J. M.

Proc. Natl. Acad. Sci USA **2006**, *103*, 11497.

- (8) (a) Dervan, P. B.; Urbach, A. R. In *Essays in Contemporary Chemistry*; Quinkert, G., Kisakürek, M. V., Eds.; Verlag Helvetica Chimica Acta: Zurich, 2000, pp 327. (b) Urbach, A. R.; Dervan, P. B. *Proc. Natl. Acad. Sci USA* **2001**, *98*, 4343. (c) Urbach, A. R.; Love, J. J.; Ross, S. A.; Dervan, P. B. *J. Mol. Biol.* **2002**, *320*, 55. (d) Marques, M. A.; Doss, R. M.; Urbach, A. R.; Dervan, P. B. *Helv. Chim. Acta* **2002**, *85*, 4485. (e) James W. Puckett, Dissertation (Chapter 2B).
- (9) Singh-Gasson, S.; Green, R. D.; Yue, Y. J.; Nelson, C.; Blattner, F.; Sussman, M. R.; Cerrina, F. *Nat. Biotechnol.* **1999**, *17*, 974.
- (10) (a) Trauger, J. W.; Baird, E. E.; Dervan, P. B. *Nature* **1996**, *382*, 559–561. (b) Trauger, J. W. Ph. D. Thesis, California Institute of Technology, 1999.
- (11) For derivation of these equations see the Supporting Information for Bulyk, M. L.; Huang, X. H.; Choo, Y.; Church, G. M. *Proc. Natl. Acad. Sci USA* **2001**, *98*, 7158.
- (12) (a) $c \times [\text{PA}]$ was 1.7×10^{-5} for polyamide **2**. (b) c was 80.2×10^3 and $[\text{PA}]$ was 5.5×10^{-9} M for polyamide **4**. (c) To view plots reflecting the same curve fits of **Figure 3.8** on a log–log scale, please see **Figure 3.9**.
- (13) Although the data for polyamide **2** (**Figure 3.8A**) maps intensity and K_a values using the linearized Eq 2, this fit is distinct from that obtained by fitting the data to a line of the form $y = mx + b$, which includes an intensity-axis intercept term. While very small in this case, the differences in the slopes and intercepts of the lines may indicate error both in the background correction of the microarray and in the DNase I footprinting data. To correct for this possibility, we propose the use of an error term, ϵ , that would modify Eqs 1 and 2 to the following:

$$\text{Intensity} = c \times \Theta + \epsilon = c \times \{K_a[\text{PA}]\}/\{1+K_a[\text{PA}]\} + \epsilon = c \times \{[\text{PA}]\}/\{K_d+[\text{PA}]\} + \epsilon$$
(Eq 1e) and $\text{Intensity} = c \times [\text{PA}]/K_d + \epsilon = c \times [\text{PA}] \times K_a + \epsilon$ (eq 2e). When fitting the intensity and K_a data for polyamide **2** to the modified Eq 2e, one finds a marginally improved fit ($R^2 = 0.97$), although the curve fit for polyamide **4** using Eq 1e is

- unimproved ($R^2 = 0.99$). For polyamide **2**, $c \times [\text{PA}] = 1.5 \times 10^{-5}$ and $\varepsilon = 5.5 \times 10^3$. For polyamide **4**, $c = 81.2 \times 10^3$, $[\text{PA}] = 5.7 \times 10^{-9}$ M, and $\varepsilon = -1.1 \times 10^3$.
- (14) For the relationship between polyamides **1** and **2**, $a = 0.0253$ and $n = 1.115$. For the relationship between polyamides **3** and **4**, $a = 349.83$ and $n = 0.637$.
- (15) Workman, C. T.; Yin, Y. T.; Corcoran, D. L.; Ideker, T.; Stormo, G. D.; Benos, P. V. *Nucleic Acids Res.* **2005**, *33*, W389.
- (16) Figure 3.11 utilized ten variable bases and contained a background GC content of 50%; Figure 3.12 utilized ten variable bases and two bases, each flanking the 5' and 3' portion of the variable region contained a background GC content of 58%. These background corrections were utilized in the motif searching parameters.
- (17) There are 1258 occurrences of a full six base pair match sequence, TTACGT. Double this number of highest intensity sequences was also searched yielding only modest changes in the data. The sequence logo is reported for the 2516 highest intensity sequences.
- (18) There are 24 occurrences of a full nine base pair match sequence, AAGAAGAAG on the microarray. Double this number of highest intensity sequences was searched in addition to searching only the 24 highest intensities, yielding only small changes in the data. The sequence logo reported contains the 48 highest intensity sequences.
- (19) To convert intensity to K_a , we have included the error term ε in our calculations. This gives modified versions of Eqs 3 and 4, $K_a = (\text{Intensity} - \varepsilon) / \{[\text{PA}] \times (c - \text{Intensity} + \varepsilon)\}$ and $K_a = (\text{Intensity} - \varepsilon) / \{[\text{PA}] \times c\}$, respectively.
- (20) Van Dyke, M. M.; Dervan, P. B. *Science* **1984**, *225*, 1122.
- (21) Kong, D.; Park, E. J.; Stephen, A. G.; Calvani, M.; Cardellina, J. H.; Monks, A.; Fisher, R. J.; Shoemaker, R. H.; Melillo, G. *Cancer Res.* **2005**, *65*, 9047.
- (22) Belitsky, J. M.; Nguyen, D. H.; Wurtz, N. R.; Dervan, P. B. *Bioorg. Med. Chem.* **2002**, *10*, 2767.
- (23) Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1996**, *118*, 6141.

- (24) Colantuoni, C.; Henry, G.; Zeger, S.; Pevsner, J. *Bioinformatics* **2002**, *18*, 1540–1541.
- (25) Bolstad, B. M.; Irizarry, R. A.; Astrand, M.; Speed, T. P. *Bioinformatics* **2003**, *19*, 185.
- (26) (a) Iverson, B. L.; Dervan, P. B. *Nucleic Acids Res* **1987**, *15*, 7823. (b) Maxam, A. M.; Gilbert, W. *Methods Enzymol* **1980**, *65*, 499.

Chapter 4

Progress Towards Polyamide Inhibition of Myc-Activated Gene Expression by Antagonism of the E-box Fragment 5'-WCGWGW-3'

The work outlined in this chapter was performed in collaboration with Daniel A. Harki (California Institute of Technology).

Abstract

Deregulated expression of the Myc transcription factor is a potent and frequently encountered defect of human cancer, and a known contributor to cancer pathogenesis or disease progression. Inactivation of the *myc* oncogene or Myc protein represents a promising avenue for the development of novel anticancer agents. We describe here the synthesis and biochemical characterization of sequence-specific, DNA-binding polyamides as antagonists of Myc-activated gene expression. Our goal is to employ DNA-bound polyamides to sterically block Myc-activated gene expression in human cancer cells. While several of these polyamides showed promising cellular uptake properties in live cells, modulation of Myc-regulated gene expression was not observed in MCF-7 (breast cancer) or A549 (lung carcinoma) cell lines. We believe that a study of polyamide inhibition of Myc-regulated genes has the potential to uncover a novel mechanism for the regulation of Myc-overexpressed genes, and we anticipate future developments in polyamide design will allow for cellular uptake in other relevant cell lines which may display a different genetic response to polyamide treatment.

4.1 Introduction

The c-Myc (Myc) transcription factor, the protein product of the *c-myc* (*myc*) proto-oncogene, participates in an array of cellular processes, such as proliferation, growth, differentiation, and apoptosis (**Figure 4.1**).^{1,2} Unfortunately, deregulated Myc expression is a potent and frequently encountered defect of human cancer, and a known contributor to cancer pathogenesis or disease progression.^{1,3} Myc is deregulated and overexpressed in most cancer cells, and de-activation of *myc* in established, Myc-induced transgenic tumors triggers proliferative arrest and redifferentiation of tumor cells, resulting in rapid tumor regression.⁴

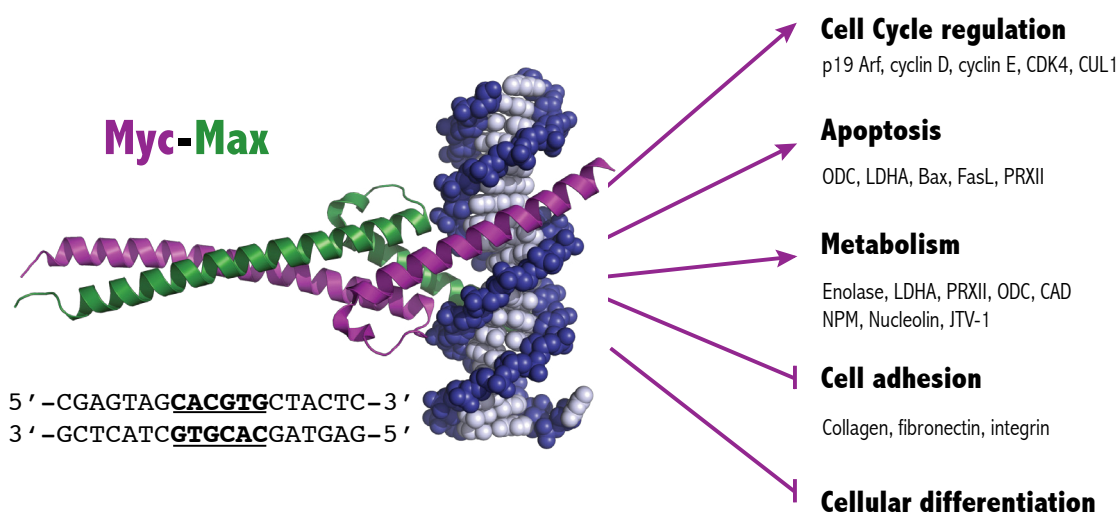


Figure 4.1 The Myc-Max transcription factor

The Myc-Max basic-helix-loop-helix leucine zipper heterodimer binds a target DNA site, known as the E-box, having a consensus sequence 5'-CACGTG-3'. c-Myc regulates downstream target genes resulting in activation of cell cycle regulation, apoptosis, or inhibition of cell adhesion. Examples of c-Myc target genes associated cellular functions are listed.

Structurally, Myc resembles most transcription factors, possessing a helix-loop-helix leucine zipper surface for protein dimerization and a basic region that forms contacts with DNA surfaces.^{3,5,6} The active Myc transcription factor exists as a heterodimer with Max (Myc associated protein X), which binds DNA and interacts with other distinct proteins.^{3,5}

Interestingly, the Myc-Max heterodimer targets different DNA sequences for gene activation and suppression. Myc-Max activates gene expression by binding E-box sequences in gene promoters (5'-CACGTG-3') and suppresses gene expression by targeting proximal promoter regions (not E-boxes) in other genes.^{1,3,7-13} Myc also associates with proteins involved in chromatin remodeling, transcriptional regulation, and the maintenance of Myc stability.¹³

Inactivation of the *myc* oncogene or Myc protein represents a promising avenue for the development of novel anticancer agents.^{3,14} Previous studies directed at the *myc* oncogene have utilized triplex-forming oligonucleotides (TFOs) that target the *myc* promoter. In particular, a TFO conjugated to daunomycin (DNA intercalator) was recently reported to reduce Myc promoter activity in prostate and breast cancer cells.¹⁵ Similarly, RNAi knockdown of *myc* expression¹⁶ and antisense oligodeoxynucleotides targeting *myc*¹⁷ in the MCF-7 breast cancer cell line have yielded promising results. Recently, small molecule inhibitors of Myc-Max heterodimerization have been identified from library screening¹⁸ and diversity-oriented synthesis,¹⁹ and shown to inhibit Myc-dependent proliferation and Myc-Max heterodimerization, respectively, in micromolar quantities.

However, in spite of these encouraging results, Myc is essential for cell growth, differentiation, proliferation and, critical for cancer patients, stem cell compartment maintenance of regenerative adult tissues such as the gastrointestinal tract, skin and bone marrow. Hence, blocking Myc function systemically might trigger devastating and irreversible side effects. As Myc functions in cells by both initiating and suppressing gene expression as the Myc-Max heterodimer, we hypothesize that cellular abatement of Myc will invariably contribute to additional gene deregulation.³ It has been estimated that Myc can bind ~25,000 sites in the human genome and regulate a significant number of genes.²⁰ Hence, decreasing the cellular levels of Myc or disrupting Myc-Max heterodimerization

may relinquish control of genes normally repressed by Myc. Furthermore, an autoregulatory pathway that monitors Myc protein levels has been proposed as the mechanism by which cells regulate Myc gene expression.²¹ Decreasing intracellular Myc may actually increase Myc expression. DNA-binding polyamides offer a new approach for controlling Myc overexpression that circumvents such problems—the blocking of Myc-Max heterodimer binding to E-box DNA sequences (5'-CACGTG-3') by a sequence-specific hairpin polyamide (**Figure 4.2**).

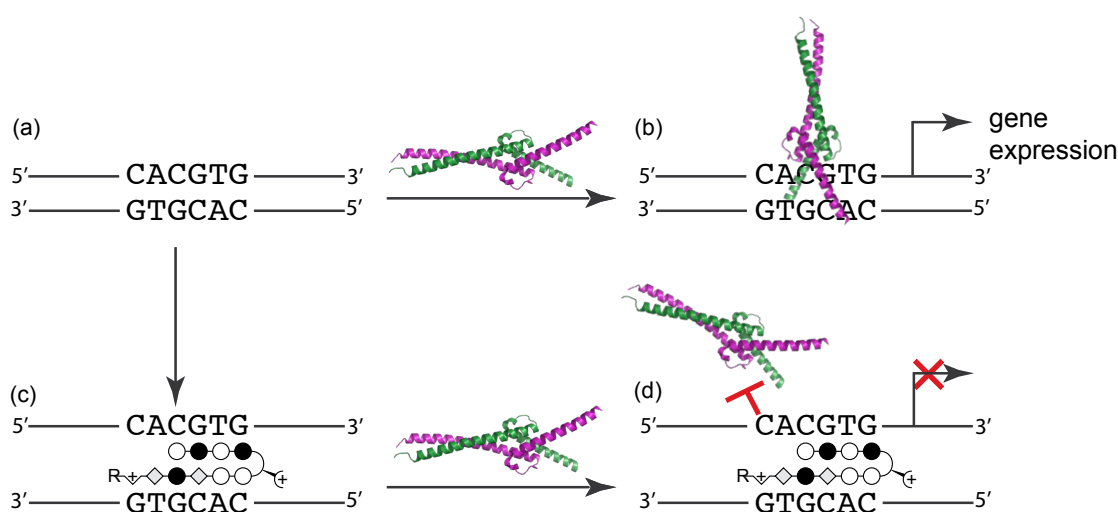


Figure 4.2 Polyamide inhibition of Myc-Max DNA binding

Simplified model of the inhibition of Myc-Max modulated gene expression by a sequence-specific DNA-binding polyamide. (a) E-box DNA sequence found in gene promoters (b) Myc-Max binds and activates gene expression (c) Polyamide bound to the E-box DNA sequence (d) Bound polyamide antagonizing Myc-Max binding and inhibiting gene expression.

We describe here the synthesis and biochemical characterization of sequence-specific, DNA-binding polyamides as antagonists of Myc-activated gene expression. Our goal is to employ DNA-bound polyamides to sterically block Myc-activated gene expression in human cancer cells. Accordingly, a small library of polyamides targeting the E-box (5'-CACGTG-3') sequence recognized by Myc in gene promoters has been synthesized and evaluated biochemically. By antagonizing Myc binding to E-box DNA sequences, as

opposed to decreasing cellular levels of Myc protein or disrupting Myc-Max heterodimer formation, polyamides may counteract the effects of elevated Myc levels in cells without relinquishing control of genes normally repressed by Myc. Importantly, this strategy should be relatively specific for inhibiting only the Myc processes associated with gene activation, since only a minority of the known Myc-Max binding sites *in vivo* possess the consensus 5'-CACGTG-3' sequence targeted by our proposed polyamide library.²⁰

4.2 Experimental Design

I. Design and synthesis of a polyamide library to bind 5'-WCGTGW-3'

A small library of polyamides targeting the Myc E-box sequence (5'-CACGTG-3') was synthesized on solid-phase using protocols developed in the Dervan laboratory.²² Compounds **1-6** possess different modifications to the polyamide skeleton to optimize binding affinity and specificity (**Figure 4.3**). The molecular design of these compounds is based upon previous studies in the Dervan laboratory that examined the utility of placing a β -alanine unit targeting the G•C base pair proximal to the C•G in the terminal position of a polyamide²³ and immediately following the chiral turn.²⁴ This initial library consists of compounds containing beta-alanine-diaminopropylamine ("Dp") tails.

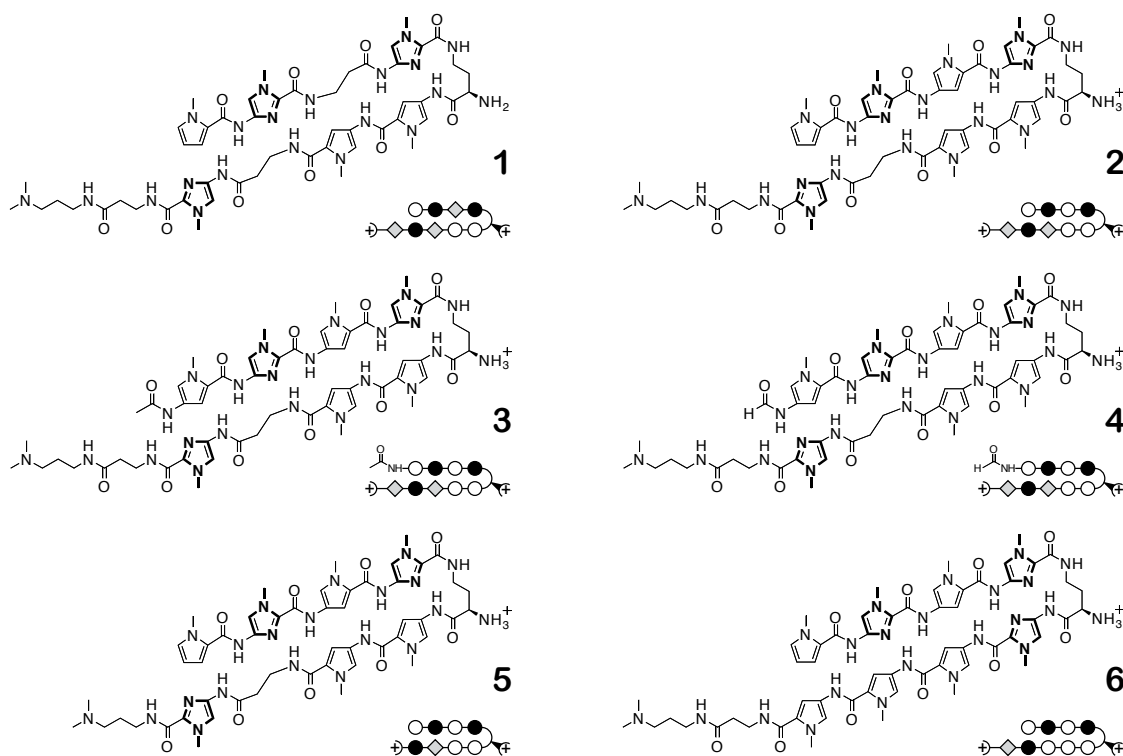


Figure 4.3 Initial polyamide library.

Compounds 1-6, synthesized by D. Harki and used in initial binding affinity screening studies

III. Design and synthesis of a polyamide library to disrupt Myc-Max DNA binding

Based on the results of the screening, a small library of polyamides derived from compounds **1** and **2** was designed to bind the 5'-WCGTGW-3' sequence of the E-box in gene promoters and block Myc-Max-activated gene expression. Over the course of experimentation, this library evolved to contain polyamides of several different scaffolds and tail substituents, including hairpins containing both diaminopropylamine ("Dp") as well as diamino-N-methyl-dipropyl amine ("triamine") tails, hairpins containing a β -alanine moiety in the tail region, hairpins lacking this β -alanine tail moiety, and even cyclic polyamides (**Figure 4.4**). Polyamides **7-11** contain an isophthalic acid tail modification, a modification that has been shown in previous studies by the Dervan laboratory to facilitate cellular uptake.²⁶

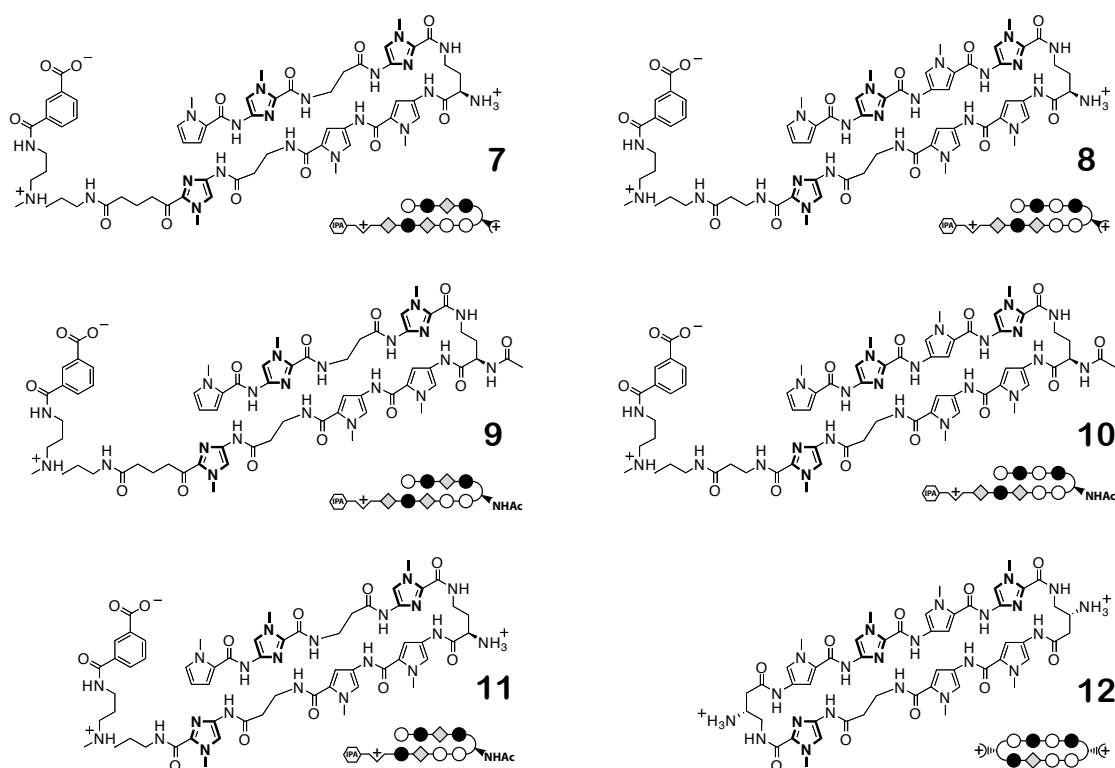


Figure 4.4 Library of compounds synthesized by K. Muzikar and D. Harki for biological studies

IV. Assessment of binding affinities and specificities of biological polyamide library

The binding affinities and specificities of polyamides **7** and **8** were rigorously characterized by quantitative DNase I footprint titrations as well as thermal melting analysis (**Figure 4.5**).²⁷ These polyamides are analogues of **1** and **2**, respectively, and contain the isophthalic acid tail modification often noted to encourage positive nuclear localization in cell culture. The 5'-terminal base pair was interrogated for specificity since this position is often deemed the hardest to obtain specificity. As shown in **Figure 4.5**, extremely high binding affinities for both polyamides were observed. Polyamide **7** bound its match DNA sequence with $K_a = 4.0 \times 10^9 \text{ M}^{-1}$ and polyamide **8** bound the same match site with $K_a = 1.3 \times 10^{10} \text{ M}^{-1}$. To our delight, high specificities against binding to the mismatch sites was observed for both compounds; with polyamide **8** exhibiting 10-fold preference for match site I versus mismatch sites II-IV. Based on these findings we concluded that polyamides **7** and **8** are more than sufficient to proceed to biological studies.

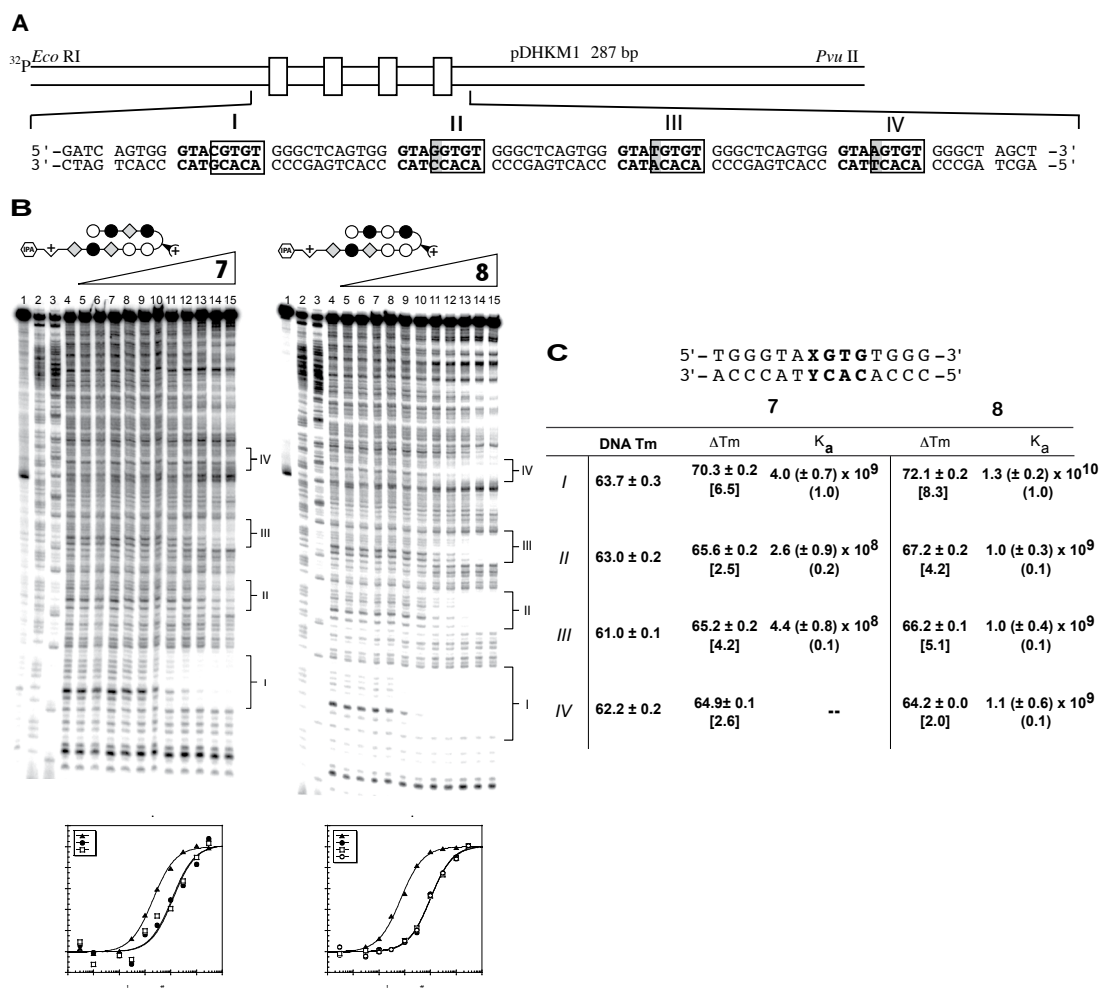


Figure 4.5 Binding studies of polyamides 7 and 8

A) Sequence of the pDHKM1 plasmid insert *B*) Storage phosphor autoradiograms from quantitative DNase I footprint titrations of polyamides 7 (left) and 8 (right). Lane 1, intact DNA, lane 2 G reaction, lane 3 A reaction, lane 4, DNase control. lanes 5-15: DNase I digestion products in the presence of 1pM, 3pM, 10pM, 30pM, 100pM, 300pM, 1nM, 3nM, 10nM, 30nM, 100nM polyamide, respectively. *C*) Summary of binding affinities as measured by DNase footprinting and by melting temperature studies of duplex DNA (2 nmoles/oligo) treated with polyamides (2.0-2.4 nmoles). Averages and S.D. were calculated from at least four analyses. Values in brackets represent ΔT_m versus untreated DNA duplex. Values in parenthesis represent K_a relative to K_a for match site. Melting temperature experiments were performed by D. Harki.

V. Evaluation of polyamide uptake and localization in cell culture

In order to assess the ability of the polyamides to traffic across the cell membrane and localize to the nucleus of cells, fluorescein isothiocyanate (FITC) conjugates **13-16** were synthesized (**Figure 4.6**) so that uptake could be evaluated by confocal microscopy.

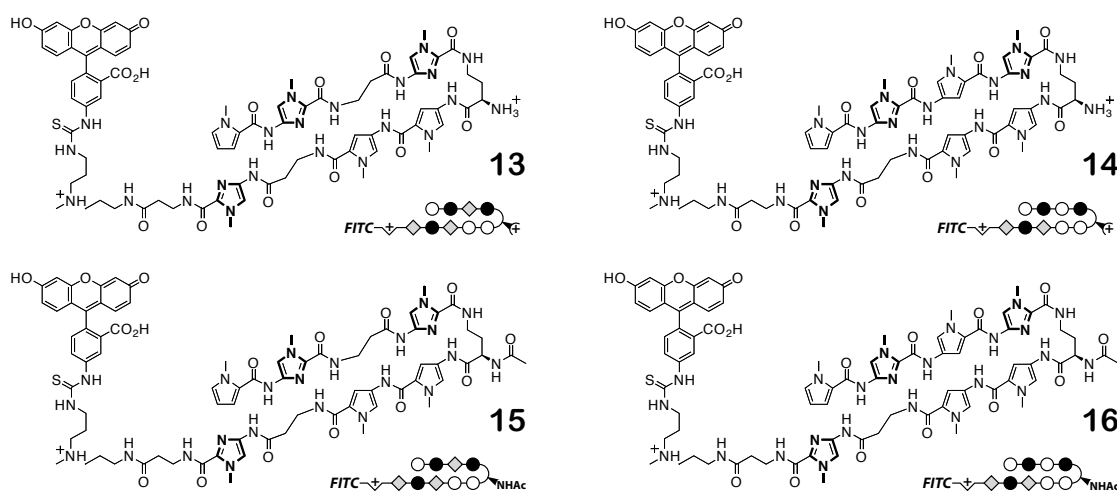
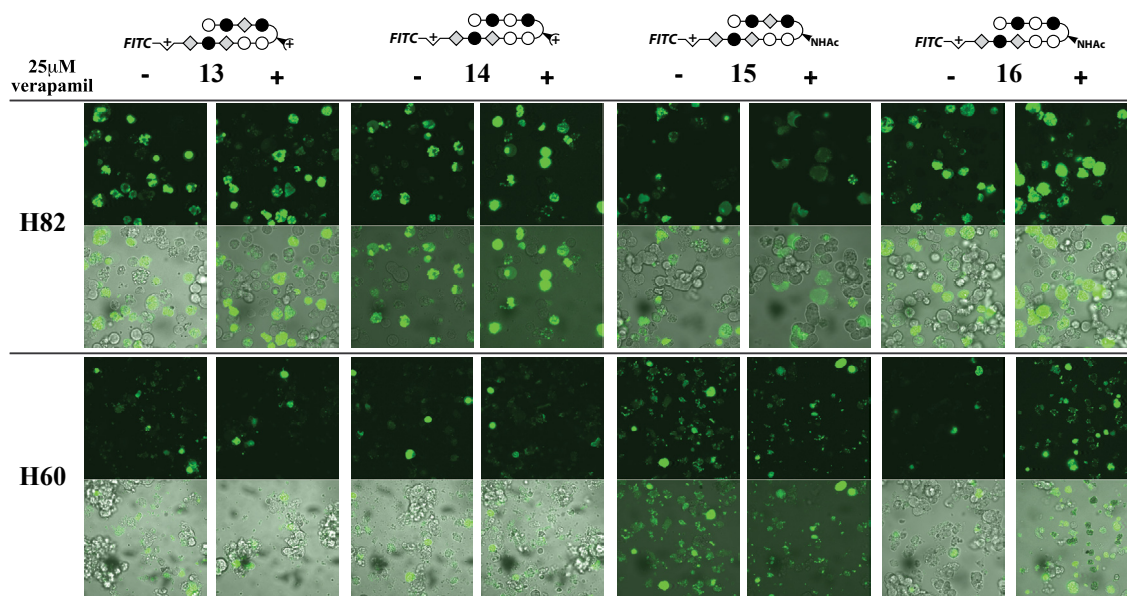


Figure 4.6 Fluorescein isothiocyanate (FITC) conjugate polyamides synthesized by K. Muzikar and D. Harki

Polyamides **13** and **14** were added at two micromolar concentration to A549 (human lung carcinoma), MCF-7 (human breast cancer), LNCaP (human prostate cancer) and P493 (human lymphocytes) cells. For A549, MCF-7, and LncCaP cell lines (**Figure 4.7**) significant amounts of fluorescence was observed in the nucleus when dosed both alone and when co-administered with the calcium channel blocker verapamil (25 μ M), suggesting that both compounds access the nucleus and accumulate to reasonable concentrations in these cell lines. In P493 cells, however, uptake is negligible for both compounds, with only small amounts of **13** and **14** localized to the nucleus when co-administered with verapamil (25 μ M).

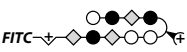
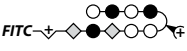
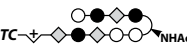

Confocal laser scanning microscopy images of polyamide-fluorescein conjugates **13** and **14** in several different cancer cell lines. For each group of pictures, the left images show 2 μ M polyamide treatment with 14 hour incubation; the right images show 2 μ M polyamide and 25 μ M verapamil treatment with 14 hour incubation; top images show fluorescence signal from polyamide; bottom images show fluorescence signal overlaid on visible light image.



*Confocal laser scanning microscopy images of polyamide-fluorescein conjugates **13-16** in NCI H82, H60 cell lines. For each group of pictures, the left images show 2μM polyamide treatment with 14 hour incubation; the right images show 2μM polyamide and 25μM verapamil treatment with 14 hour incubation; top images show fluorescence signal from polyamide; bottom images show fluorescence signal overlaid on visible light image.*

The favorable uptake of **13** and **14** in the aforementioned three cell lines suggests both polyamides have cell permeability properties sufficient for further *in vivo* studies, while studies in P493 cells were not pursued. Additionally, polyamides **13-16** were added at 2 μ M concentration to NCI-H60 (non-small-cell lung cancer) and NCI-H82 (small-cell lung cancer) cells (**Figure 4.8**), which are known to have elevated levels of c-myc gene.²⁸ Uptake of all four compounds in these cell lines appears favorable both with and without verapamil co-administration, although less so in the NCI-H60 cells. **Table 4.2** summarizes the uptake properties of compounds **13-14** in the cell lines described above.

Table 4.2 Cellular localization of polyamide–dye conjugates in cultured cells

	25mM verapamil	P493	MCF-7	A549	LnCap	H82	H60
	-	-	++	++	++	++	+
13	+	++	++	++	++	++	+
	-	-	++	++	++	++	+
14	+	++	++	++	++	++	+
	-	<i>n.d.</i>	<i>n.d.</i>	<i>n.d.</i>	<i>n.d.</i>	+	++
15	+	<i>n.d.</i>	<i>n.d.</i>	<i>n.d.</i>	<i>n.d.</i>	+	++
	-	<i>n.d.</i>	<i>n.d.</i>	<i>n.d.</i>	<i>n.d.</i>	++	-
16	+	<i>n.d.</i>	<i>n.d.</i>	<i>n.d.</i>	<i>n.d.</i>	++	++

++, Nuclear staining exceeds that of the medium; +, nuclear staining less than or equal to that of the medium, but still prominent; -, very little nuclear staining, with the most fluorescence seen in the cytoplasm and/or medium; --, no nuclear staining. *n.d.* = not determined.

VI. Analysis and selection of Myc-inducible gene targets

While it is estimated that ~25,000 genes are regulated in some way by Myc, there is a surprising dearth of known primary gene targets of Myc. Identification of Myc-regulated genes has generally relied on experimental activation of Myc followed by monitoring of changes in mRNA levels.^{29,30} More than 10 investigative works reported the use of high-throughput screening based on cDNA microarrays or the SAGE assay, significantly expanding the list of genes that are up- or downregulated by Myc. Based on an updated online compilation,³¹ this list now includes over 1500 genes. It remains unclear, however, how many of these genes are direct targets of Myc. In many studies a large fraction of Myc-target genes respond weakly, or even fail to respond, to Myc activation, depending on the cell type or experimental conditions used. Additionally, lists of genes identified in high-throughput screens such as RNA microarrays,³² ChIP-screening,^{33,34} or ChIP-PET³⁵ are only partially overlapping, especially between cell lines, and many genes were identified only once. Thus, we still possess a fragmentary picture of the loci that are directly targeted by Myc, and no accurate estimate of their numbers.

Table 4.3 E-box and surrounding DNA sequences of several known direct Myc target genes

gene	E-box sequence	gene	E-box sequence
RCC1	GCCGGCCC <u>CACGTG</u> AAGCCCGGAGG	ECA39	CTGAGCGC <u>CACGTG</u> TCACTGCACAG
RCC1	ACTTCGAC <u>CACGTG</u> TGACTTGTGTG	cdc25A	GGGCCTGC <u>CACGTG</u> CACCCCGCCCCG
ODC	TGTGCGGC <u>CACGTG</u> TCGCGAGGCCC	cdc25A	ACTACAGA <u>CACGTG</u> CCACCACACCC
ODC	GCAGGGGA <u>CACGTG</u> GTCGCCGAGCG	PT	GCAACGAG <u>CACGTG</u> GCCTGGGGCGC
eIF4e	CCATCGGC <u>CACGTG</u> ACCAGTCCTTT	P53	TCCCCTCC <u>CACGTG</u> CTCACCCTGGC
eIF4e	ATATCCGT <u>CACGTG</u> GCCAGAAGCTG	MrDb	TGACTCAC <u>CACGTG</u> CATACTATGG

Shown in bold is the six base pair E-box site. Underlined are 5'-WCGWGW-3' polyamide match sites. Sites without an underline present a GC base pair mismatch under the turn moiety.

Despite this complexity, we have identified a small list of genes that are known direct targets of Myc, and whose promoter E-boxes have been rigorously identified (**Table 4.3**). We anticipate that polyamides that bind and interfere with the expression of these genes would be useful in then mining the genome for other, perhaps previously unidentified, direct targets of Myc. Of the genes identified in **Table 4.3**, the gene that has by far the most relevance to cancer biology is the cap-binding protein eukaryotic translation initiation factor 4E (eIF4E). A functional promoter region of eIF4E has been identified 400 nucleotides upstream of the transcription initiation site with two essential E-box motifs in the immediate promoter region.³⁸ Of these two E-boxes, only one has a full match site (5'-ACGTGA-3') for the polyamide, while the other (5'-ACGTGG-3') presents a G•C base pair under the turn moiety (**Figure 4.9**)

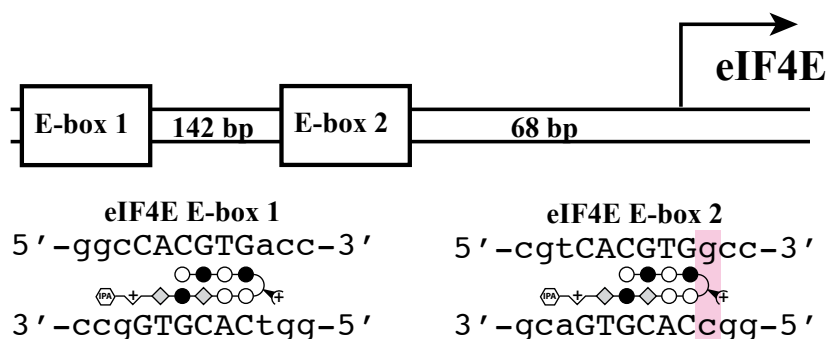


Figure 4.9 Schematic representation of E-box locations relative to transcription start site of eIF4E

Shown are the sequences of the two E-boxes with a representative polyamide shown bound. Note that while there is a full polyamide match binding site in E-box 1, there is a severe mismatch under the turn moiety of the polyamide, it is likely this polyamide has a poor binding affinity for this site.


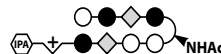
During initiation of transcription, eIF4E binds the 7-methylguanosine cap of mRNA and recruits a transcript to the translational machinery. Increased cap-dependent mRNA translation rates are frequently observed in human cancers. Mechanistically, many human tumors often overexpress the cap binding protein eukaryotic translation initiation factor 4E (eIF4E), leading to enhanced translation of numerous tumor-promoting genes.

eIF4E-specific antisense oligonucleotides have been shown to repress expression of eIF4E-regulated proteins (e.g., VEGF, cyclin D1, survivin, c-myc) in human cells (HeLa, A549, MCF-736) as well as induce apoptosis of the carcinogenic cells. Intravenous administration of eIF4E-specific antisense oligonucleotides has been shown to reduce eIF4E expression in human tumor xenografts, as well as significantly suppress tumor growth.³⁷

VII. Assessment of binding affinities of polyamides towards eIF4E E-box

The initial thermal melting temperature studies were performed on 12-mer oligos that were designed to have an unambiguous polyamide binding site (5'-TACGTGT-3') in order to probe the specificity of the polyamide in the 5' position avoiding the palindromic sequence 5'-CACGTG-3'. However, it is the second sequence that is relevant within the context of the eIF4E E-box, hence we decided to perform a T_m analysis on a 12-mer oligo containing the sequence 5'-CACGTG-3'. **Table 4.4** shows that polyamides **7** and **11** demonstrate essentially the same thermal stabilization energy ($\Delta T_m \sim 4^\circ\text{C}$) as each other for the eIF4E duplex. While comparisons cannot be drawn relating ΔT_m s across different DNA oliomers, this is the same general magnitude of stabilization previously seen on the 5'-TACGTGT-3' sequence.

Table 4.4 Thermal melting temperature studies of polyamides **7** and **11** on eIF4e E-box 1

T _m oligo:	5' - GGC CACGTG ACC - 3' 3' - CCG GTCAC TGG - 5'	5' - TGGG TACGTGT GGG - 3' 3' - ACCC ATGCAC ACC - 5'
DNA only	62.4 ± 0.6	63.7 ± 0.3
7 	65.6 ± 0.5 [3.2]	70.3 ± 0.2 [6.5]
11 	66.4 ± 0.5 [4.0]	n.d.

Melting temperature studies of duplex DNA (2 nmoles/oligo) treated with polyamides (2.0-2.4 nmoles). Averages and S.D. were calculated from at least four analyses. Values in brackets represent ΔT_m versus untreated DNA duplex. n.d. = not determined.

VIII. Selection of inducible and endogenously Myc-overexpressed cell lines

Unlike previous transcription factors that have been inhibited by polyamide-DNA binding, the Myc-Max transcription factor does not function in response to an antagonist; that is, this is not an endogenously inducible system. This gives an added level of complexity to the assessment of gene regulatory activity, hence for our initial studies of these polyamides in cell culture we intended to use an artificial system that would allow us to control cellular Myc levels. Two such cell lines were available from the literature, and we requested and received these cell lines.

P-493 cells:

The P493-6 cell line is a model for cell cycle activation by *myc* in Burkitt lymphoma cells. P493-6 is a human EBV-EBNA1 positive B-cell line in which *myc* is expressed under the control of a tetracycline-regulated promoter.³⁹ These cells were sent as a gift from Professor Dirk Eick (Institute of Clinical Molecular Biology and Tumor Genetics, Munich, Germany). Unfortunately, as can be seen in **Figure 4.7**, cellular localization of compounds **13** and **14** was negligible in this cell line and further studies were not pursued.

MCF-7cl35 cells:

The Shiu lab at University of Manitoba has developed a clonal MCF-7 human breast cancer cell line (MCF-7cl35) harboring a stably-transfected human c-myc gene, whose expression is controlled by the bacterial reverse tetracycline transcription activator protein.⁴⁰ Expression of endogenous genomic c-myc gene in MCF-7 cells is abolished by the potent pure estrogen antagonist, ICI 182,780. These cells were sent as a gift from Professor Robert P.C. Shiu (University of Manitoba, Winnipeg, Manitoba, Canada) and were cultured as instructed by the Shiu lab, which is as reported in their papers: in DMEM with sodium pyruvate, glutamine (4mM), penicillin-streptomycin (4mM), and 5% FBS.⁴

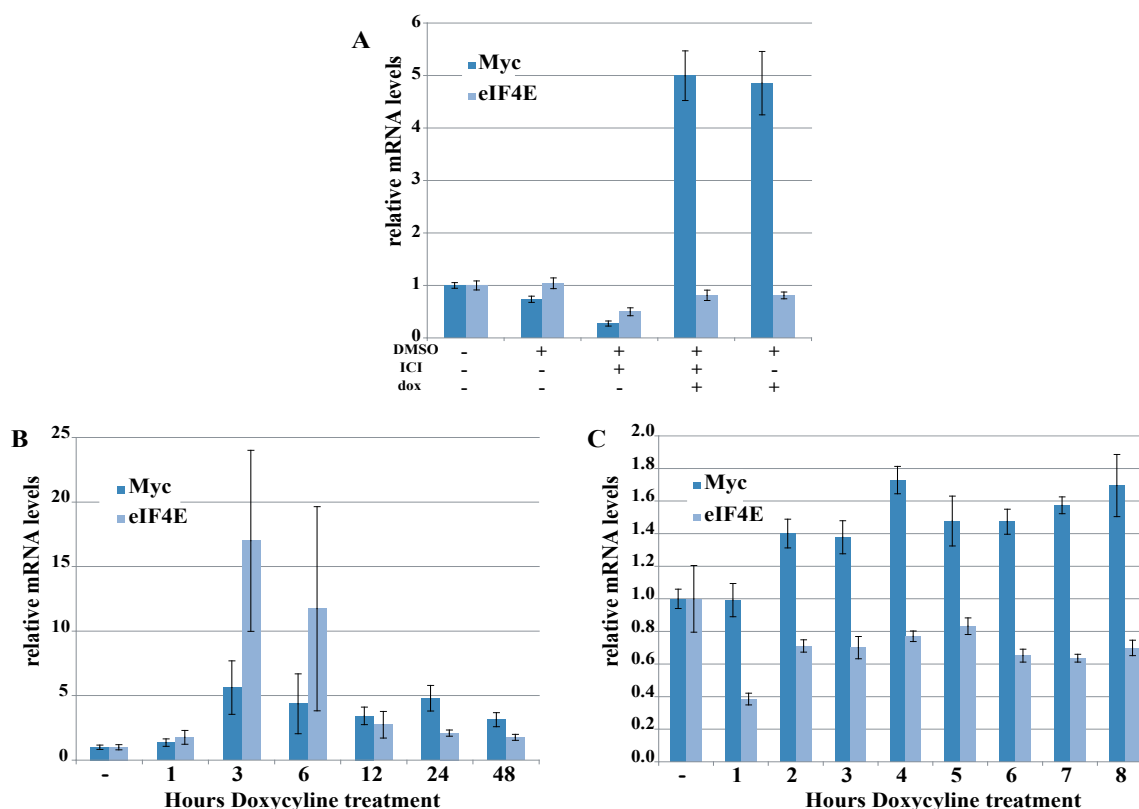


Figure 4.10 Doxycycline-induced Myc and eIF4E expression in MCF7-cl35 cells

A) Myc expression is induced 5-Fold by 12-hour incubation with doxycycline and and repressed by ICI inhibitor compound in accordance with literature precedent (cell passage number 3). B) Doxycycline time course indicates Myc is induced maximally 5-fold at 3 hours doxycycline treatment, while eIF4E is induced over 15-fold at this time point (cell passage number 4). C) Doxycycline treatment fails to induce Myc or eIF4E over a 1-8 hour range (cell passage number 7).

Myc induction by doxycycline (a water-soluble version of tetracycline) as well as repression of endogenous *myc* expression by ICI was reproduced when the cell line was first received. However, as the cells were passaged over time, the intensity of the *myc* signal induction decreased from 5-fold (passage number three) to less than 2-fold (passage number seven). Additionally, while Myc expression was induced, no induced expression of the target gene eIF4E was seen in most instances, and what induction is seen for both genes is wildly variable. **Figure 4.10** shows a compilation of data that represents the irreproducible nature of data garnered from this cell line. The large error bars seen in **Figure 4.10** were not

a one-time occurrence, but were seen in every experiment performed (over 10 independent experiments). We hypothesize that the plasmid containing the inducible tet-Myc promoter may not be as stably integrated as we expected. There is no additive in the culture medium as described by the Shiu group to keep a selective pressure on the cells to maintain the plasmid, and it may be that the plasmid does not remain in the cells after several passages. While working with an inducible Myc cell line was desirable, we decided to move our efforts into an endogenous system.

NCI-H82 cells:

NCI-H82 is a small-cell lung cancer cell line in which c-myc DNA sequences are amplified about 25-fold, with a resulting 24-fold increase in c-myc RNA relative to normal cells.⁴¹ This line grows as aggregates of cells in suspension, and is available for purchase from the American Type Culture Collection (ATCC). In order to establish a baseline for comparison for gene repression, Validated Stealth siRNA was purchased from Invitrogen and optimized for use with the NCI-H82 cell line. **Figure 4.11** shows a representative qRT-PCR experiment using the siRNA against Myc in NCI-H82 cells.

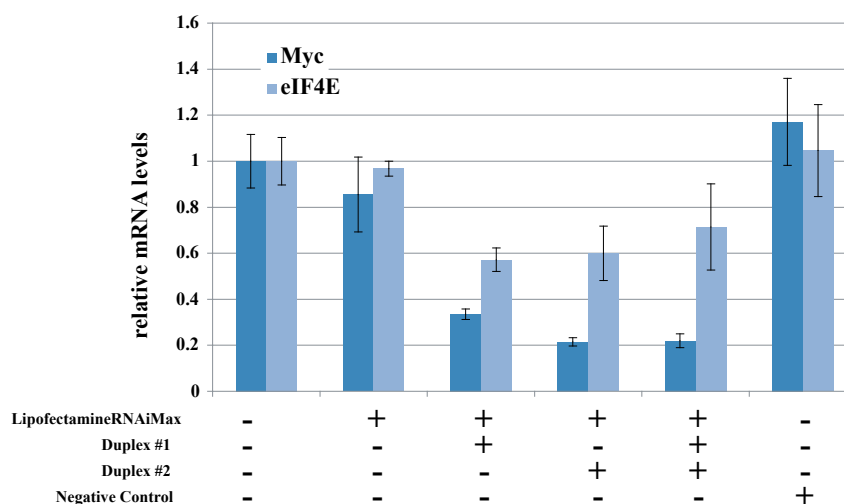


Figure 4.11 siRNA against Myc in NCI-H82 cells

IX. qRT-PCR analysis of Myc-inducible gene expression following polyamide treatment

NCI-H82 cells were dosed with compounds **7-10** at 10 μ M concentration for 48 hours, after which time their mRNA was harvested, reverse transcribed, and analyzed by quantitative real-time polymerase chain reaction (qRT-PCR) for the expression levels of Myc and eIF4E along with several other genes from **Table 4.3** (ODC, RCC1). Unfortunately, no appreciable effect on eIF4E gene expression (or ODC or RCC1) is seen following 1, 5, or 10 μ M concentration treatment of compounds **7-10** (data not shown). **Figure 4.12** shows representative results of three independent biological trials, while error bars indicate the SD of four technical biological replicates of the given trial. While the levels of the actual Myc gene are slightly reduced, no appreciable effect is seen on eIF4E expression in excess of that which may be caused by the slight Myc repression. siRNA treatment concurrent with these biological trials indicates that reduced levels of expression would have likely been detectable if present.

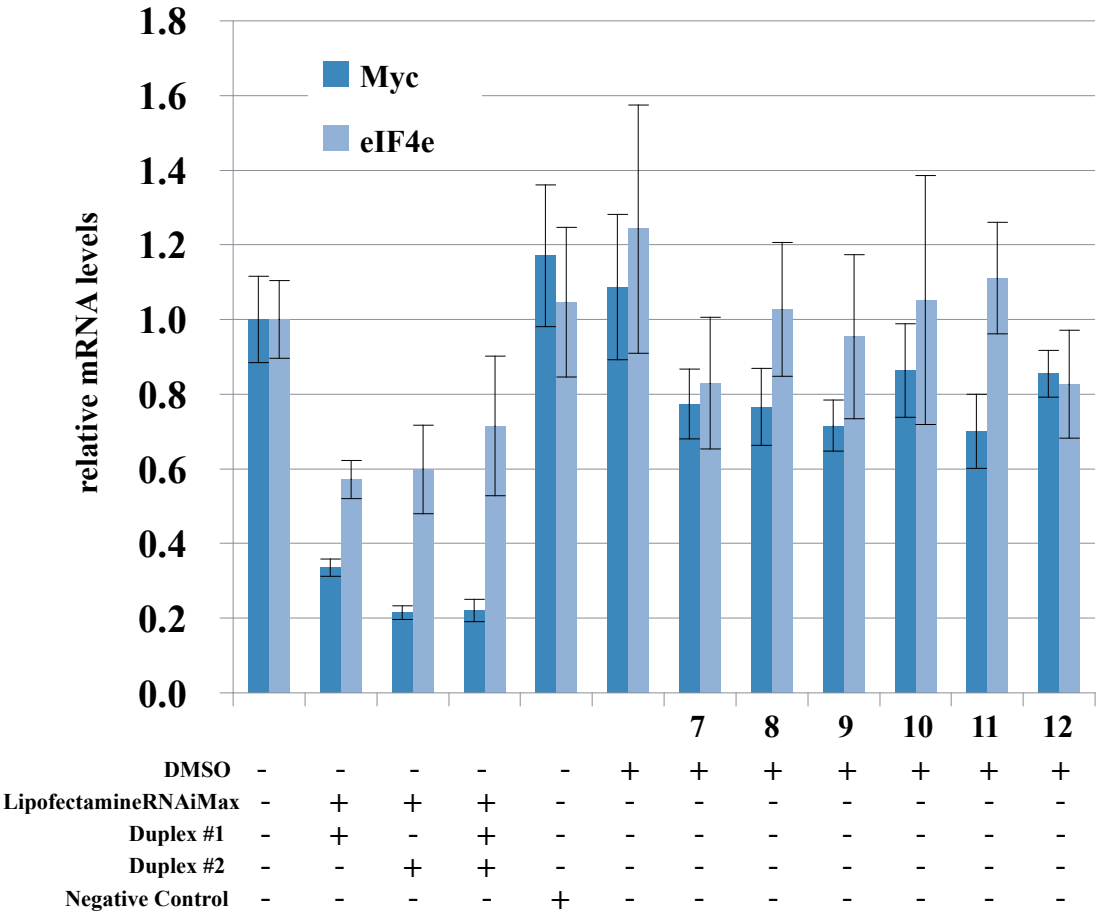


Figure 4.12 qRT-PCR results in NCI-H82 cells

Representative qRT-PCR results after 48 hours treatment of siRNA or 10 μ M concentrations of polyamides 7-12 in NCI-H82 cells. Error bars represent SD of at least three technical replicates.

X. Enzyme-linked immunosorbent assay (ELISA)

In collaboration with Summer Undergraduate Research Fellow (SURF) Nickolaus Krall, an ELISA assay was developed that would allow us to measure the protein concentration of the eIF4e and Myc proteins. In the first step an ELISA^{42,43} was developed to measure eIF4E levels in crude cell lysate. Starting from a design suggested previously⁴⁴ the optimal concentrations of antibodies, buffer compositions, and appropriate incubation conditions were determined. To test the hypothesis that the polyamides only affect Myc binding to

E-box DNA and hence expression levels of target genes but not Myc protein expression itself an ELISA for the measurement of Myc protein levels in crude cell lysate was also developed. Due to lack of gene response to polyamide treatment, these assays were never used to measure change of protein levels in response to polyamide, but the detailed experimental protocol developed can be found in the Materials and Methods section of this chapter (Section 4.4, ix).

4.3 Closing Remarks

We have described the development of cell-permeable sequence-specific polyamides that bind to the Myc-Max E-box DNA sequence. By antagonizing the discrete binding event between Myc-Max and its cognate DNA sequence we hope to regain control of deregulated gene expression facilitated by Myc overexpression. A library of matched and mismatched polyamides was prepared by solid- and solution-phase synthesis, and their binding affinities and specificities for E-box (5'-CACGTG-3') DNA was determined by quantitative DNase I footprinting titration and thermal melting temperature analysis. The favorable cellular permeability and intracellular localization properties of the FITC conjugates of the compounds was observed by laser-scanning confocal microscopy. However, the ability of the polyamides to inhibit Myc-Max binding *in vivo* was not demonstrated in the cell lines and conditions described here. It is possible that improvements to the uptake of these or other E-box-targeted polyamides into alternate cell lines would demonstrate different biological effects in those cell lines. We believe that a study of polyamide inhibition of Myc-regulated genes has the potential to uncover a novel mechanism for the regulation of MYC-overexpressed genes, and we anticipate future developments in polyamide design may give this project new direction.

4.4 Materials and Methods

i. Polyamide synthesis

Chemicals not otherwise specified are from Sigma-Aldrich.

Synthesis of polyamides 1-6:

Polyamides **1-6** were synthesized by Dan Harki using pre-loaded Boc- β -Ala-PAM resin (50 mg, 0.81 meq/g, Peptides International) according to published manual solid-phase synthesis protocols.^{22,45} The resin was cleaved with neat 3-dimethylamino-1-propylamine (1 mL) at 37 °C for 16 h. Products were purified by preparatory reverse-phase HPLC on a Beckman Gold system using either a Waters Delta-Pak 25°—100 mm, 15 μ m 300 Å C18 PrepPak Cartridge reverse-phase column or a Varian Dynamax 21.4°—250 mm Microsorb 8 μ m 300 Å C8 reverse-phase column in 0.1% (w/v) TFA with acetonitrile as the eluent. The appropriate fractions were lyophilized after characterization by analytical HPLC, UV-visible spectroscopy, and MALDI-TOF or ESI mass spectrometry.

1: (MALDI-TOF) [M+H]⁺ calcd for C₅₁H₇₀N₂₁O₁₀⁺ 1136.6, observed 1136.6

2: (MALDI-TOF) [M+H]⁺ calcd for C₅₄H₇₁N₂₂O₁₀⁺ 1187.6, observed 1187.6

3: (MALDI-TOF) [M+H]⁺ calcd for C₅₆H₇₄N₂₃O₁₁⁺ 1244.6, observed 1244.6

4: (MALDI-TOF) [M+H]⁺ calcd for C₅₅H₇₂N₂₃O₁₁⁺ 1230.6, observed 1230.7

5: (MALDI-TOF) [M+H]⁺ calcd for C₅₁H₆₆N₂₁O₉⁺ 1116.5, observed 1116.5

6: (MALDI-TOF) [M+H]⁺ calcd for C₅₇H₇₂N₂₃O₁₀⁺ 1238.6, observed 1238.6

Synthesis of polyamides 7-10:

Polyamides **7-10** were synthesized by Katy Muzikar and Dan Harki using Boc- β -Ala-PAM resin according to published manual solid-phase synthesis protocols

as referenced above. The protected FmocHN- γ -turn amine was deprotected with 20% piperidine in DMF and reprotected as the Boc derivative with a solution of Boc₂O (Fluka) and DIEA in DMF. The Boc-protected resin was cleaved with 1 mL of 3,3'-diamino-N-methyldipropylamine (triamine) at 37 °C for 16 h. Products were purified by preparatory reverse-phase HPLC and the appropriate fractions were lyophilized after characterization by analytical HPLC, UV-visible spectroscopy, and MALDI-TOF or ESI mass spectrometry. Conjugates were formed by pre-activating isophthalic acid (3.0 equiv) with PyBOP (2.9 equiv, Novabiochem) in a solution of DIEA (20 equiv) and DMF at 37 °C for 30 min, followed by reaction of the activated solution with the polyamide for 1 h at room temperature. Conjugates were deprotected with neat TFA for 10 min at room temperature before purification by preparatory reverse-phase HPLC. Lyophilization of the appropriate fractions yielded the polyamide conjugates **7-10**, which were characterized as described above. Extinction coefficients were calculated according to standard protocols.²⁷ (i.e., estimating 69,500 cm⁻¹ M⁻¹ for standard 8-ring polyamides leads to an approximation of 8,690 cm⁻¹ M⁻¹ per ring for polyamides containing a β -alanine moiety in place of a ring. Calculated extinction coefficient was rounded to the nearest hundred, thus the extinction coefficient for a 7-ring polyamide is calculated to be 60,800 cm⁻¹ M⁻¹).

7: (MALDI-TOF) [M+H]⁺ calcd for C₅₁H₇₀N₂₁O₁₀⁺ 1136.6, observed 1136.6

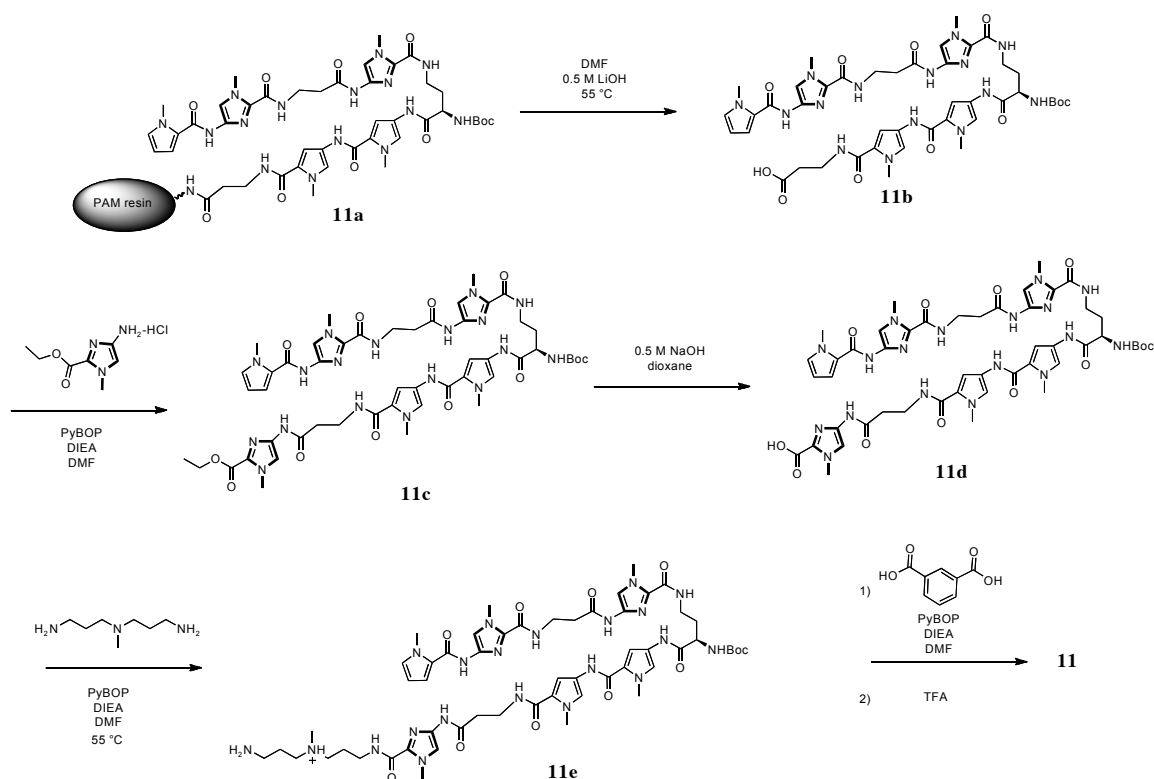
8: (MALDI-TOF) [M+H]⁺ calcd for C₅₄H₇₁N₂₂O₁₀⁺ 1187.6, observed 1187.6

9: (MALDI-TOF) [M+H]⁺ calcd for C₅₆H₇₄N₂₃O₁₁⁺ 1244.6, observed 1244.6

10: (MALDI-TOF) [M+H]⁺ calcd for C₆₃H₈₀N₂₂O₁₄⁺ 1369.6, observed 1369.9

Synthesis of polyamide 11:

Polyamide **11** was synthesized by a different route than the rest of the compounds owing to the fact that imidazole monomer does not stably load onto oxime resin, and synthesis on Boc- β -Ala-Pam resin installs the undesired β -alanine moiety.



Scheme 4.1 Synthesis of polyamide **11**.

As shown in **Scheme 4.1**, this necessitated several off-resin synthetic steps. Because a β -alanine moiety was desired in the second position from the resin, the core of **11** was synthesized according to standard protocols as described above on Boc- β -Ala-Pam resin. The core **11a** was saponified off the resin via incubation with 0.5M lithium hydroxide (LiOH) to provide the carboxylic acid **11b**, which was purified by preparatory reverse-phase HPLC. This was subsequently coupled with 4-[(tert-Butoxycarbonyl)amino]-1-methylimidazole-2-carboxylic acid (NH₂-Im-OEt) following established protocols,⁵³ yielding compound **11c**, which was purified by preparatory reverse-phase HPLC. **11c** was saponified in 0.5 M sodium hydroxide (NaOH) to yield the carboxylic acid **11d** which was purified by preparatory reverse-phase HPLC. This was conjugated to 3,3'-diamino-N-methyldipropylamine via PyBOP activation, resulting in compound **11e**, which was purified by preparatory reverse-phase HPLC.

11e was conjugated to IPA, deprotected, and purified by HPLC as described above.

11b: (MALDI-TOF) $[M+H]^+$ calcd for $C_{43}H_{56}N_{15}O_{11}^+$ 957.4, observed 958.5

11c: (MALDI-TOF) $[M+H]^+$ calcd for $C_{50}H_{65}N_{18}O_{12}^+$ 1109.5, observed 1109.6

11d: (MALDI-TOF) $[M+H]^+$ calcd for $C_{48}H_{61}N_{18}O_{12}^+$ 1081.4, observed 1037.7
 $([M-CO_2+H]^+)$

11e: (MALDI-TOF) $[M+H]^+$ calcd for $C_{55}H_{79}N_{21}O_{11}^+$ 1208.6, observed 1208.8

11: (MALDI-TOF) $[M+H]^+$ calcd for $C_{58}H_{74}N_{21}O_{12}^+$ 1256.6, observed 1256.8

Synthesis of polyamide 12:

Cyclic polyamide **12** was synthesized following the methodology established for solution-phase cyclic polyamide synthesis.⁵⁴ Briefly, polyamide half-strands **12a** and **12b** (**Scheme 4.2**) were synthesized and were conjugated together by solution-phase synthetic methods.⁵³ After conjugation of the half-strands, **12c** was saponified to **12d**, which was converted to the pentafluorophenol (Pfp) ester **12e**. Each step of this synthesis up to this point proceeded in similar yields to those previously published and all intermediates were characterized by ¹HNMR and analyzed for purity by analytical scale reverse-phase HPLC. Each product was >90% pure as assessed by these methods. Pfp ester **12e** was deprotected and cyclized to form **12f**, an insoluble compound that was not characterized. The Cbz protecting groups present in **12f** were cleaved and the resulting solution purified by reverse-phase HPLC to yield compound **12** in trace amounts. This synthesis provided the 30 nmols necessary to perform qRT-PCR analysis of its effect in cell culture for one biological replicate. Due to the lack of biological activity of cyclic polyamide **12**, the detailed characterization of these 34 intermediates has been excluded from this thesis.

12: (MALDI-TOF) $[M+H]^+$ calcd for $C_{58}H_{74}N_{21}O_{12}^+$ 1256.6, observed 1256.8

Synthesis of polyamides 13-16:

Polyamides **13-16** were synthesized by Katy Muzikar and Dan Harki using Boc- β -Ala-PAM resin according to published manual solid-phase synthesis protocols as referenced above. The protected FmocHN- γ -turn amine was deprotected with 20% piperidine in DMF and reprotected as the Boc derivative with a solution of Boc₂O (Fluka) and DIEA in DMF. The Boc-protected resin was cleaved with 1 mL of 3,3'-diamino-N-methyldipropylamine (triamine) at 37 °C for 16 h. Products were purified by preparatory reverse-phase HPLC and the appropriate fractions were lyophilized after characterization by analytical HPLC, UV-visible spectroscopy, and MALDI-TOF or ESI mass spectrometry.

Conjugates were formed by reacting fluorescein-5-isothiocyanate (FITC, Invitrogen) with the polyamide in a solution of DIEA (20 equiv) and DMF for 1 h at room temperature. Conjugates were deprotected with neat TFA for 10 min at room temperature before purification by preparatory reverse-phase HPLC. Lyophilization of the appropriate fractions yielded the polyamide conjugates **13-16**, which were characterized as described above. Extinction coefficients were calculated according to standard protocols.

13: (MALDI-TOF) $[M+H]^+$ calcd for $C_{74}H_{86}N_{23}O_{15}S^+$ 1568.6, observed 1569.7

14: (MALDI-TOF) $[M+H]^+$ calcd for $C_{77}H_{87}N_{24}O_{15}S^+$ 1619.7, observed 1620.8

15: (MALDI-TOF) $[M+H]^+$ calcd for $C_{76}H_{88}N_{23}O_{16}S^+$ 1610.7, observed 1611.6

16: (MALDI-TOF) $[M+H]^+$ calcd for $C_{79}H_{89}N_{24}O_{16}S^+$ 1661.7, observed 1662.6

ii. UV absorption spectrophotometry of DNA thermal stabilization

Melting temperature analysis was performed on a Varian Cary 100 spectrophotometer equipped with a thermo-controlled cell holder possessing a cell path length of 1 cm. An aqueous solution of 10 mM sodium cacodylate, 10 mM KCl, 10 mM $MgCl_2$, and 5 mM $CaCl_2$ at pH 7.0 was used as analysis buffer. This buffer was degassed under vacuum and DNA duplexes and polyamides were added to a final concentration of 2 μ M DNA, 3 μ M polyamide for each experiment. Prior to analysis, samples were heated to 90 °C and cooled to a starting temperature of 23 °C with a heating rate of 5 °C/min for each ramp. Denaturation profiles were recorded at $\lambda = 260$ nm from 23 to 90 °C with a heating rate of 0.5 °C/min. Each sample was subjected to denaturation in technical duplicate, and the reported numbers are the average of at least three experimental replicates. The reported melting temperatures were defined as the maximum of the first derivative of the denaturation profile.

iii. Plasmid preparation

Plasmid pDHKM1 was constructed following standard protocols. The following hybridized insert (Integrated DNA Technologies) was ligated into the BamHI/HindIII polycloning site in pUC19 Plasmid (Sigma) using a Rapid DNA Ligation Kit (Roche) according to the manufacturer's protocols:

5'-GATC AGTGG GTACGTGT GGGCTCAGTGG GTAGGTGT GGGCTCAGTGG
GTATGTGT GGGCTCAGTGG GTAAGTGT GGGCT-3'

5'-AGCT AGCCC ACACTTAC CCACTGAGCCC ACACATAC CCACTGAGCCC
ACACCTAC CCACTGAGCCC ACACGTAC CCACT-3'

The ligated plasmid was then transformed into JM109 subcompetent cells (Promega) by standard methods (30 minute incubation on ice followed by 45 second heat shock (42 °C) followed by 1 hour incubation at 37 °C). Colonies were selected for α -complementation on agar plates containing 50 mg/L ampicillin, 120 mg/L IPTG, and 40 mg/L X-gal after overnight growth at 37 °C. Cells were harvested after 16 h growth at 37 °C in LB medium containing 50 mg/L ampicillin. Plasmid was then purified by midi-prep kit (Promega). The presence of the desired inserts was determined by capillary electrophoresis dideoxy sequencing methods (Laragen).

iv. Preparation of 5'-labeled DNA for DNase I footprinting

Two primer oligonucleotides, 5'-AATTCGAGCTCGGTACCCGGG-3' (forward, corresponding to EcoRI restriction enzyme cut site) and 5'-CTGGCACGACAGGTTTCCCGA-3' (reverse, corresponding to the PvuII restriction

enzyme cut site) were constructed for PCR amplification. The forward primer was radiolabeled using [α - 32 P]-dATP (MP Biomedicals) and polynucleotide kinase (Roche), followed by purification using ProbeQuant G-50 spin columns (GE Healthcare). The desired PCR product was generated from the plasmid pDHKM1 using the primer pair and Expand High Fidelity PCR Core Kit (Roche) following the manufacturer's protocol. The labeled fragment was purified on a 7% nondenaturing preparatory polyacrylamide gel (5% cross-link) and visualized by autoradiography. The radiolabeled band was excised, crushed, and soaked overnight (14 hours) in 2 M NaCl. The gel pieces were removed by centrifugal filtration and the DNA was precipitated with 2-propanol (1.5 volumes). The pellet was washed with 75% ethanol, lyophilized to dryness, and then resuspended in 1 mL of RNase-free water. Chemical sequencing reactions were performed according to published protocols.^{46,47}

v. DNase I footprint titrations

All reactions were carried out in a volume of 400 μ L according to published procedures.²⁷ Footprinting was performed on 1 pM, 3 pM, 10 pM, 30 pM, 100 pM, 300 pM, 1 nM, 3 nM, 10 nM, 30 nM, 100 nM solutions of polyamides **7** and **8**, where polyamide solutions were quantitated at $\lambda=310$ nm using $\epsilon=52,100$ M⁻¹ cm⁻¹ (compound **7**) and $\epsilon=60,800$ M⁻¹ cm⁻¹ (compound **8**). Polyamides were equilibrated with the radiolabeled DNA for 14 h prior to DNase I cleavage at 23 °C. DNA was precipitated and subjected to gel electrophoresis as previously described. Quantitation by storage phosphor autoradiography and determination of equilibrium association constants were as previously described.²⁷

vi. Cell culture

The human lung carcinoma cell line NCI-H82 (American Type Culture Collection HTB-175) was maintained in RPMI-1640 Medium (ATCC Catalog No. 30-2001) supplemented with L-glutamine (4 mM), penicillin/streptomycin (4 mM) and 10% fetal bovine serum (Irvine Scientific). Cell growth and morphology were monitored by phase-contrast microscopy.

The human lung carcinoma cell line A549 (ATCC CCL-185) was maintained in F-12K Medium, (ATCC Catalog No. 30-2004) supplemented with L-glutamine (4 mM), penicillin/streptomycin (4 mM) and 10% fetal bovine serum as recommended by the ATCC. Cell growth and morphology were monitored by phase-contrast microscopy.

The human breast adenocarcinoma cell line MCF-7 (American Type Culture Collection HTB-22) was maintained in Eagle's Minimum Essential Medium (ATCC Catalog No. 30-2003), supplemented with 0.01 mg/ml bovine insulin; 10% fetal bovine serum, as recommended by the ATCC. Cell growth and morphology were monitored by phase-contrast microscopy.

The human prostate adenocarcinoma cell line LnCaP (ATCC CRL-1740) was maintained in RPMI-1640 Medium (ATCC Catalog No. 30-2001), supplemented with 10% fetal bovine serum, as recommended by the ATCC. Cell growth and morphology were monitored by phase-contrast microscopy.

vii. Confocal microscopy

Cells were trypsinized for 5-10 min at 37°C, centrifuged for 5 min at 2,000 rpm and 5 °C in a Beckman-Coulter Allegra 6R centrifuge, and resuspended in fresh medium to a concentration of 1×10^6 cells per mL. Incubations were performed by adding 150 μ L of cells into culture dishes equipped with glass bottoms for direct imaging (MatTek, Ashland, MA). The cells were grown in the glass-bottom culture dishes for 24 h. Then 5 μ L of a 60 μ M polyamide solution was added and the cells. For verapamil-treated cells, 5 μ L 750mM verapamil (\pm) was added in addition to polyamide. The plates were then incubated in a 5% CO₂ atmosphere at 37 °C for 10-14 h. Imaging was performed with a 40x oil-immersion objective lens on a Zeiss LSM 5 Pascal inverted laser scanning microscope. Polyamide–fluorescein conjugate fluorescence and visible-light images were obtained using standard filter sets for fluorescein.^{46,47} 12-Bit images were analyzed using Zeiss LSM and ImageJ software.

viii. Determination of relative mRNA levels via qRT-PCR

siRNA:

Myc Validated Stealth DuoPak SKU# 12936-50, was purchased from Invitrogen. This included Duplex #1 and Duplex #2, of proprietary sequence and concentration. Stealth RNAi™ siRNA Negative Control Kit (Invitrogen, Cat. No. 12935-100) was purchased and the “Medium GC Duplex #1” was used. siRNA was transfected using LipofectamineRNAiMAX (Invitrogen, Cat. No. 13778-075) according to the manufacturer’s instructions. The amounts of these reagents used in most assays was 1 μ L Lipofectamine RNAiMax and 1.4 μ L of supplied siRNA.

Note: at the time of the printing of this thesis, Invitrogen has discontinued the siRNA product described. Brief attempts to use the replacement product yielded poor results, but a thorough optimization of conditions was not performed.

RNA isolation:

Cells were plated in 24-well dishes at a density of 40×10^4 cells/mL in 0.5 mL of culture medium without antibiotic supplements (which are lethal in combination with siRNA treatment) and allowed to attach for 16-20 hours. Polyamides or siRNA were added (according to instructions above) and the cells were incubated for 48 hours. The medium was removed, cells were washed with ice-cold PBS and immediately lysed with RLT buffer from an RNeasy kit (Qiagen). Further RNA isolation was carried out with the RNeasy kit as described in the manufacturer's manual. The isolated total RNA was quantified. The yields were 12-15 μ g per well.

Reverse transcription:

A 2.5 μ g sample of total RNA was used to reverse-transcribe cDNA using Superscript II reverse transcriptase (Invitrogen) according to the manufacturer's protocol. Random hexamers were used as primers. The total volume for each RT reaction was 20 μ L.

Real-time quantitative RT-PCR:

Analysis was performed using the Myc and eIF4E gene primers described below, purchased from Integrated DNA Technologies. Quantitative real-time RTPCR was performed using Applied Biosystems SYBR Green RT-PCR master mix according to the manufacturer's instructions. Temperature cycling and detection of the SYBR Green emission were performed with an ABI 7300 real-time instrument using Applied Biosystems Sequence Detection System version 1.2. Statistical analysis was performed on three independent experiments.

To amplify the 88-bp fragment from the 3'-translated region of Myc:

Forward primer: 5'-CAGCGACTCTGAGGAGGAAC-3'

Reverse primer: 5'-CTCTGACCTTTTGCCAGGAG -3'

To amplify the 118-bp fragment from the 3'-translated region of eIF4E:

Forward primer: 5'-TTTTGGGCTCTGTACAACCA-3'

Reverse primer: 5'-CTCCCCGTTTGTTTTTCTCA -3'

RNA was standardized by quantification of the β -glucuronidase gene as an endogenous control,⁴⁸ Using the following primers:

Forward primer: 5'-CTCATT TGGAATTTTGCCGATT -3'

Reverse primer: 5'- CCGAGTGAAGATCCCCTTTTTA -3'

ix. Determination of relative protein levels via Enzyme-Linked Immunosorbent Assay (ELISA)

In collaboration with Nickolaus Krall, Summer Undergraduate Research Fellow (SURF) student, summer 2008:

Cells were incubated at 37 °C, 5% CO₂. Cells were induced by the addition of 2 μ l of 1mg/ml doxycyclin (in H₂O) to each well at the indicated time point for each experiment. At the indicated harvest time point for each experiment the cells were detached from the plates by the addition of trypsin. Cells were counted via hemocytometer and subsequently pelleted by centrifugation at 130 RCF for 10 minutes. The supernatant was removed and the pellet was rinsed twice with PBS and repelleted before the addition of RIPA buffer supplemented with protease inhibitor cocktail (PIC) and PMSF. Pellets were agitated by

pipetting and left on ice for 20-30 minutes before cell debris was pelleted by centrifugation at 1000 RCF for 5 minutes. The supernatant was transferred to a fresh tube and frozen at -80°C in an isopropanol slow-freeze bath.

Normalization

Frozen cell lysate was thawed on ice and adjusted to a constant protein or DNA concentration prior to determining eIF4E concentrations. Total protein concentration was determined with a Bradford assay.¹⁹ One ml of Bradford reagent (BioRAD) was added to 10 µL of crude lysate and 10 µl of water contained in a 1 ml plastic cuvette and mixed vigorously. A ready-to-go BSA standard set (BioRAD) was used as a reference. One ml of Bradford reagent was added to 10 µl of BSA standard solution and 10 µl of RIPA buffer contained in a 1 ml plastic cuvette. The absorbance was measured at 595 nm (A_{595}) with a UV/Vis Spectrometer (Agilent) and a linear standard curve constructed from the BSA reference dilutions. Using this standard curve total protein concentrations in samples were calculated from the A_{595} . All measurements were carried out in duplicate and the average value used to adjust all samples to the same total protein concentration by dilution with appropriate amounts of RIPA buffer. For normalization to constant total DNA, DNA was extracted from crude cell lysate using a PCR purification kit (Qiagen). DNA content was measured with the built in function of a Nano Drop UV/Vis spectrophotometer and samples adjusted to the same DNA content with RIPA buffer.

eIF4E ELISA:

96-well plates (Greiner BioOne Microton 600) were coated overnight with donkey anti-mouse IgG (Jackson, 50µl of 5µg/ml antibody in PBS with 0.05% NaN₃ per well, 2-8°C). Plates were subsequently washed (3x 200µl of PBS with 0.05% Tween20 per well), then blocked (200 µl of 5% dry milk powder in PBS per well) for 1 h at room temperature. Plates were washed again, then incubated with mouse anti-eIF4E IgG (BD

Signal Transduction Laboratories, 50 μ l of a 1:450 dilution in antibody buffer [i.e., PBS, 0.25% BSA and 0.05% Tween20] per well) for 1 h at room temperature. After another wash they were incubated with samples (50 μ l per well) and standards (Globozymes recombinant eIF4E; 50 μ l of 200, 100, 50, 25, 15, 10, 5, 1 and 0.1 ng/ml in antibody buffer per well) for 1 h at room temperature. Plates were then washed again and incubated with rabbit anti eIF4E (Cell Signaling, 50 μ l of a 1:1000 dilution in antibody buffer) for 1 h at room temperature, followed by another wash step. Plates were subsequently incubated with goat anti-rabbit IgG HRP conjugate (Jackson, 50 μ l of a 1:80,000 dilution in antibody buffer per well for 1 h at room temperature), washed, and finally developed with TMB substrate (Sigma, 50 μ l per well) After 30 minutes the reaction was stopped with an equal volume of 1 M HCl. The output signal was read as the absorbance at 450 nm (A450) with a Perkin Elmer multiwell plate reader. The absorbance of eIF4E standards was fit to a 4-parameter logistic curve^{17,18} using Kaleida Graph and used to calculate unknown eIF4E concentrations in samples from A450 values.

Myc ELISA:

The Myc ELISA closely followed the eIF4E ELISA protocol above but differed in choice and concentration of antibodies. Coating with donkey anti-IgG was carried out as described above. The secondary coating antibody was a monoclonal mouse anti-c-Myc IgG (Sigma, clone 9E10, 50 μ l of a 1:300 dilution in antibody buffer per well). Incubation with samples was as above. The primary detection antibody was a polyclonal antibody rabbit anti c-Myc IgG raised against the 262 N-terminal amino acids (Santa Cruz, 50 μ l of a 1:250 dilution in antibody buffer per well). The secondary detection antibody was a HRP conjugated goat anti-rabbit IgG (Jackson, 50 μ l of a 1:20,000 dilution in antibody buffer).

4.5 References

- (1) Chung, H.-J.; Levens, D. *Mol. Cells* **2005**, *20*, 157.
- (2) Schwab, M.; SpringerLink (Online service); Springer Berlin Heidelberg: Berlin, Heidelberg, 2009.
- (3) Ponzielli, R.; Katz, S.; Barsyte-Lovejoy, D.; Penn, L. Z. *Eur. J. Cancer* **2005**, *41*, 2485.
- (4) Soucek, L.; Whitfield, J.; Martins, C. P.; Finch, A. J.; Murphy, D. J.; Sodik, N. M.; Karnezis, A. N.; Swigart, L. B.; Nasi, S.; Evan, G. I. *Nature* **2008**, *455*, 679.
- (5) Fieber, W.; Schneider, M. L.; Matt, T.; Krautler, B.; Konrat, R.; Bister, K. *J. Mol. Biol.* **2001**, *307*, 1395.
- (6) Nair, S. K.; Burley, S. K. *Cell* **2003**, *112*, 193.
- (7) Barsyte-Lovejoy, D.; Mao, D. Y. L.; Penn, L. Z. *Oncogene* **2004**, *23*, 3481.
- (8) Seoane, J.; Le, H. V.; Massague, J. *Nature* **2002**, *419*, 729.
- (9) Izumi, H.; Molander, C.; Penn, L. Z.; Ishisaki, A.; Kohno, K.; Funa, K. *J. Cell Sci.* **2001**, *114*, 1533.
- (10) Staller, P.; Peukert, K.; Kiermaier, A.; Seoane, J.; Lukas, J.; Karsunky, H.; Moroy, T.; Bartek, J.; Massague, J.; Hanel, F.; Eilers, M. *Nat. Cell Biol.* **2001**, *3*, 392.
- (11) Oster, S. K.; Marhin, W. W.; Asker, C.; Facchini, L. M.; Dion, P. A.; Funa, K.; Post, M.; Sedivy, J. M.; Penn, L. Z. *Mol. Cell. Biol.* **2000**, *20*, 6768.
- (12) Dang, C. V. *Mol. Cell. Biol.* **1999**, *19*, 1.
- (13) Ponzielli, R.; Katz, S.; Barsyte-Lovejoy, D.; Penn, L. Z. *Eur. J. Cancer* **2005**, *41*, 2485.
- (14) Shachaf, C. M.; Kopelman, A. M.; Arvanitis, C.; Karlsson, A.; Beer, S.; Mandl, S.; Bachmann, M. H.; Borowsky, A. D.; Ruebner, B.; Cardiff, R. D.; Yang, Q. W.; Bishop, J. M.; Contag, C. H.; Felsher, D. W. *Nature* **2004**, *431*, 1112.

- (15) Carbone, G. M.; McGuffie, E.; Napoli, S.; Flanagan, C. E.; Dembech, C.; Negri, U.; Arcamone, F.; Capobianco, M. L.; Catapano, C. V. *Nucleic Acids Res.* **2004**, *32*, 2396.
- (16) Wang, Y. H.; Liu, S.; Zhang, G.; Zhou, C. Q.; Zhu, H. X.; Zhou, X. B.; Quan, L. P.; Bai, J. F.; Xu, N. Z. *Breast Cancer Res.* **2005**, *7*, R220.
- (17) Carroll, J. S.; Swarbrick, A.; Musgrove, E. A.; Sutherland, R. L. *Cancer Res* **2002**, *62*, 3126.
- (18) Kiessling, A.; Sperl, B.; Hollis, A.; Eick, D.; Berg, T. *Chem. Biol.* **2006**, *13*, 745.
- (19) Jemal, A.; Siegel, R.; Ward, E.; Murray, T.; Xu, J. Q.; Smigal, C.; Thun, M. J. *CA Cancer J. Clin.* **2006**, *56*, 106.
- (20) Adhikary, S.; Eilers, M. *Nature Rev. Mol. Cell Biol.* **2005**, *6*, 635.
- (21) Kim, M. K. H.; Carroll, W. L. *Cancer* **2004**, *101*, 2106.
- (22) Belitsky, J. M.; Nguyen, D. H.; Wurtz, N. R.; Dervan, P. B. *Bioorg. Med. Chem.* **2002**, *10*, 2767.
- (23) Wang, C. C. C.; Ellervik, U.; Dervan, P. B. *Bioorg. Med. Chem.* **2001**, *9*, 653.
- (24) Wang, C. C. *Ph.D. Dissertation, California Institute of Technology* **2002**.
- (25) Pilch, D. S.; Poklar, N.; Gelfand, C. A.; Law, S. M.; Breslauer, K. J.; Baird, E. E.; Dervan, P. B. *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 8306.
- (26) Nickols, N. G.; Jacobs, C. S.; Farkas, M. E.; Dervan, P. B. *Nucleic Acids Res.* **2007**, *35*, 363.
- (27) Trauger, J. W.; Dervan, P. B. *Drug-Nucleic Acid Interactions* **2001**, *340*, 450.
- (28) Bernasconi, N. L.; Wormhoudt, T. A.; Laird-Offringa, I. A. *Am. J. Respir. Cell Mol. Biol.* **2000**, *23*, 560.
- (29) Grandori, C.; Cowley, S. M.; James, L. P.; Eisenman, R. N. *Annu. Rev. Cell Dev. Bi.* **2000**, *16*, 653.
- (30) Oster, S. K.; Ho, C. S.; Soucie, E. L.; Penn, L. Z. *Adv Cancer Res* **2002**, *84*, 81.

- (31) Zeller, K. I.; Jegga, A. G.; Aronow, B. J.; O'Donnell, K. A.; Dang, C. V. *Genome Biol.* **2003**, *4*, R69.
- (32) Watson, J. D.; Oster, S. K.; Shago, M.; Khosravi, F.; Penn, L. Z. *J. Biol. Chem.* **2002**, *277*, 36921.
- (33) Fernandez, P. C.; Frank, S. R.; Wang, L. Q.; Schroeder, M.; Liu, S. X.; Greene, J.; Cocito, A.; Amati, B. *Genes Dev.* **2003**, *17*, 1115.
- (34) Li, Z. R.; Van Calcar, S.; Qu, C. X.; Cavenee, W. K.; Zhang, M. Q.; Ren, B. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 8164.
- (35) Zeller, K. I.; Zhao, X. D.; Lee, C. W. H.; Chiu, K. P.; Yao, F.; Yustein, J. T.; Ooi, H. S.; Orlov, Y. L.; Shahab, A.; Yong, H. C.; Fu, Y. T.; Weng, Z. P.; Kuznetsov, V. A.; Sung, W. K.; Ruan, Y. J.; Dang, C. V.; Wei, C. L. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 17834.
- (36) Dong, K.; Wang, R.; Wang, X.; Lin, F.; Shen, J. J.; Gao, P.; Zhang, H. Z. *Breast Cancer Res. Tr.* **2009**, *113*, 443.
- (37) Graff, J. R.; Konicek, B. W.; Vincent, T. M.; Lynch, R. L.; Monteith, D.; Weir, S. N.; Schwier, P.; Capen, A.; Goode, R. L.; Dowless, M. S.; Chen, Y.; Zhang, H.; Sissons, S.; Cox, K.; McNulty, A. M.; Parsons, S. H.; Wang, T.; Sams, L.; Geeganage, S.; Douglass, L. E.; Neubauer, B. L.; Dean, N. M.; Blanchard, K.; Shou, J.; Stancato, L. F.; Carter, J. H.; Marcusson, E. G. *J. Clin. Invest.* **2007**, *117*, 2638.
- (38) Jones, R. M.; Branda, J.; Johnston, K. A.; Polymenis, M.; Gadd, M.; Rustgi, A.; Callanan, L.; Schmidt, E. V. *Mol. Cell. Biol.* **1996**, *16*, 4754.
- (39) Schuhmacher, M.; Kohlhuber, F.; Holzel, M.; Kaiser, C.; Burtscher, H.; Jarsch, M.; Bornkamm, G. W.; Laux, G.; Polack, A.; Weidle, U. H.; Eick, D. *Nucleic Acids Res.* **2001**, *29*, 397.
- (40) Venditti, M.; Iwasiow, B.; Orr, F. W.; Shiu, R. P. C. *Int. J. Cancer* **2002**, *99*, 35.
- (41) American Tissue Culture Collection (ATCC), www.atcc.com

- (42) Hornbeck, P.; Winston, S. E.; Fuller, S. A. *Curr. Protoc. Mol. Biol.* **2001**, *Chapter 11*, Unit 112.
- (43) Sigma Aldrich, http://www.sigmaaldrich.com/Area_of_Interest/Life_Science/Antibody_Explorer/Procedures/ELISA.html (2008)
- (44) Graff, J. R.; Konicek, B. W.; Vincent, T. M.; Lynch, R. L.; Monteith, D.; Weir, S. N.; Schwier, P.; Capen, A.; Goode, R. L.; Dowless, M. S.; Chen, Y.; Zhang, H.; Sissons, S.; Cox, K.; McNulty, A. M.; Parsons, S. H.; Wang, T.; Sams, L.; Geeganage, S.; Douglass, L. E.; Neubauer, B. L.; Dean, N. M.; Blanchard, K.; Shou, J.; Stancato, L. F.; Carter, J. H.; Marcusson, E. G. *J. Clin. Invest.* **2007**, *117*, 2638.
- (45) Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1996**, *118*, 6141.
- (46) Iverson, B. L.; Dervan, P. B. *Nucleic Acids Res.* **1987**, *15*, 7823.
- (47) Maxam, A. M.; Gilbert, W. *Methods Enzymol.* **1980**, *65*, 499.
- (48) Best, T. P.; Edelson, B. S.; Nickols, N. G.; Dervan, P. B. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 12063.
- (49) Edelson, B. S.; Best, T. P.; Olenyuk, B.; Nickols, N. G.; Doss, R. M.; Foister, S.; Heckel, A.; Dervan, P. B. *Nucleic Acids Res.* **2004**, *32*, 2802.
- (50) Aerts, J. L.; Gonzales, M. I.; Topalian, S. L. *Biotechniques* **2004**, *36*, 84.
- (51) Perlstein, M. T.; Maciel, R. J. *Clin. Chem.* **1984**, *30*, 960.
- (52) Plikaytis, B. D.; Turner, S. H.; Gheesling, L. L.; Carlone, G. M. *J Clin Microbiol* **1991**, *29*, 1439.
- (53) Chenoweth, D.M.; Harki, D.A.; Dervan, P.B. *J. Am. Chem. Soc* **2009** *131*, 7175.
- (54) Chenoweth, D.M.; Harki, D.A.; Phillips, J.W.; Dose, C.; Dervan, P.B. **2009** *J. Am. Chem. Soc.*, *131*, 7182.

Appendix A

Progress Towards Incorporation of Furan Rings into Pyrrole-Imidazole Polyamides

The work outlined in this chapter was performed in collaboration with Daniel A. Gubler (California Institute of Technology).

A.1 Introduction

Lipophilicity plays a role in the pharmacokinetic (PK) properties of lead structures for animal studies, including absorption, distribution, clearance, and metabolism. Recent *in vitro* ADMET analysis of a hairpin Py/Im polyamide revealed high levels of plasma protein binding characteristic of a highly lipophilic compound.¹ As our research program using Py-Im polyamides moves from cell culture into small animal models, new monomers with comparable DNA-binding affinities to *N*-methylpyrrole (Py) and *N*-methylimidazole (Im) and decreased lipophilicity/increased water solubility may be desirable. The development and incorporation of such monomers into Py-Im polyamides would allow the ability to fine-tune the desired lipophilicity of a given polyamide by selectively mixing and matching rings.

In this regard, the recently discovered proximicin A-C natural products have drawn our attention due to their unique 4-aminofuran-2-carboxylic acid ring system and overall similarity to DNA minor groove-binding natural products netropsin (**1**) and distamycin (**2**) (**Figure A.1**). Proximicins A-C (**3-5**) possess dual amide-linked furan amino acids and an *N*-terminal methyl carbamate functionality. *C*-terminal functionalities consisting of an amide (proximicin A, **3**), tyramine (proximicin B, **4**), or tryptamine (proximicin C, **5**) characterize each member of the proximicin family. Evaluations of **3-5** against multiple cancer cell lines revealed promising antitumor activities, with GI₅₀ values in the low micromolar range against gastric adenocarcinoma (AGS), hepatocellular carcinoma (HepG2), and breast adenocarcinoma (MCF7) cell lines. These activities were significantly better than netropsin, distamycin, and analogous compounds to **3-5** that possess Py monomers in place of both furans.^{2,3}

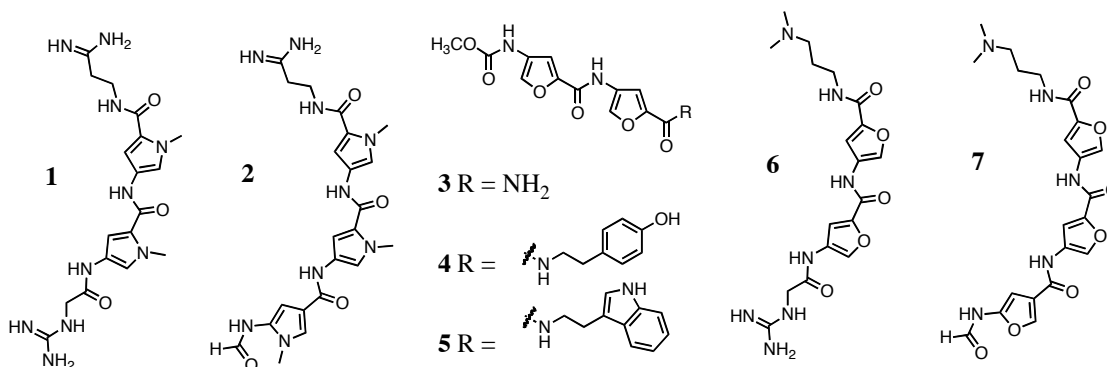


Figure A.1 DNA-binding molecules with furan heterocycle constituents

Structures of netropsin (**1**), distamycin (**2**), proximicin A-C (**3-5**), and furan analogues of netropsin (**6**) and distamycin (**7**)

The furan amino acid is identical in shape to Py with the exception of the methylated ring nitrogen, which is replaced by oxygen, and this replacement is predicted to increase the water solubility of polyamide oligomers. Incorporation of the furan (Fn) monomer into polyamides designed to inhibit gene regulation *in vivo* may provide a polyamide that demonstrates a better set of physical properties with regard to efficacy in small animal models. Since the *N*-methyl nitrogen of Py units within a polyamide are positioned away from the H-bonding edges of the aromatic amino acid ring pairs, we expect that oxygen at this position will be tolerated in terms of minor-groove binding as well as sequence recognition.

It is speculated the electronics and overall polarity of the furan amino acid in comparison to Py contributes to their enhanced antitumor activities, thus there is even stronger reasoning to evaluate this heterocycle as a potential Py replacement in hairpin polyamides. Recent efforts by Süssmuth to incorporate the furan heterocycle into netropsin and distamycin analogues produced compounds whose binding affinity for DNA has been considerably compromised, as assessed by thermal DNA denaturation stabilization experiments.³ Süssmuth suggests that the lower stabilizing effect might be due to altered

stacking interactions resulting from the different electronic structures of the two heterocyclic cores. An additional suggestion for this result is that the N-methyl pyrrole of netropsin or distamycin is bulky and nonpolar within the proximity of the anionic backbone of the DNA whereas the furan ring's free electron pair in this same position may lead to a repulsive interaction with the DNA backbone. We anticipate that these unfavorable interactions of the furan ring will be mitigated by incorporation within the context of a Py-Im polyamide, allowing the furan rings to increase polyamide solubility while the Im and Py rings maintain high DNA-binding character.

A.2 Experimental Design

I. Design and partial synthesis of a Py-Im-Fn polyamide library to bind 5'-WGGWCW-3'

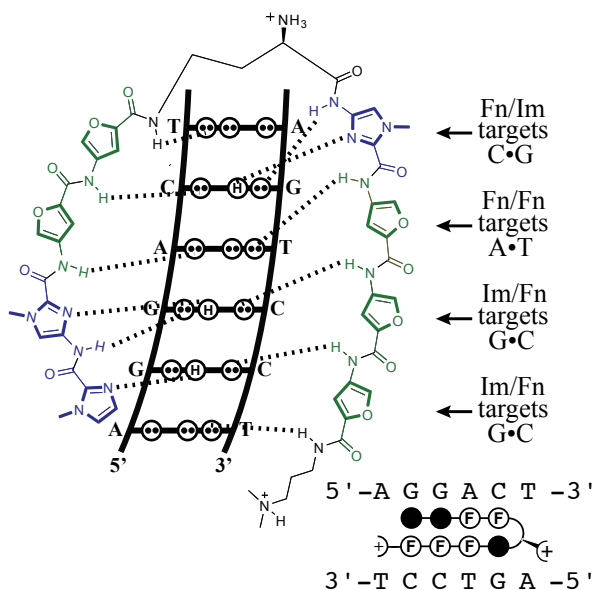


Figure A.2 Furan monomers incorporated into 5'-WGGWCW-3'targeted polyamide scaffold

We designed a library of polyamides targeting the sequence 5'-WGGWCW-3' (where W = A or T) (Figure). This scaffold was chosen because it allows for the assessment of several characteristics of the furan amino acid within a series of analogs (**Figure A.2**). As shown in **Figure A.3**, by altering which position of the parent Py-Im polyamide is altered to a furan we can assess the unit's specificity in all combinations of ring pairs: Im/Fn (**2**), Py/Fn (**3**), Fn/Im (**4**), Fn/Py (**5**) and Fn/Fn (**6**). Also, by incrementally increasing the number of furan rings incorporated into the polyamide (**7-9**), we can assess the combinatorial effect of multiple rings on lipophilicity.

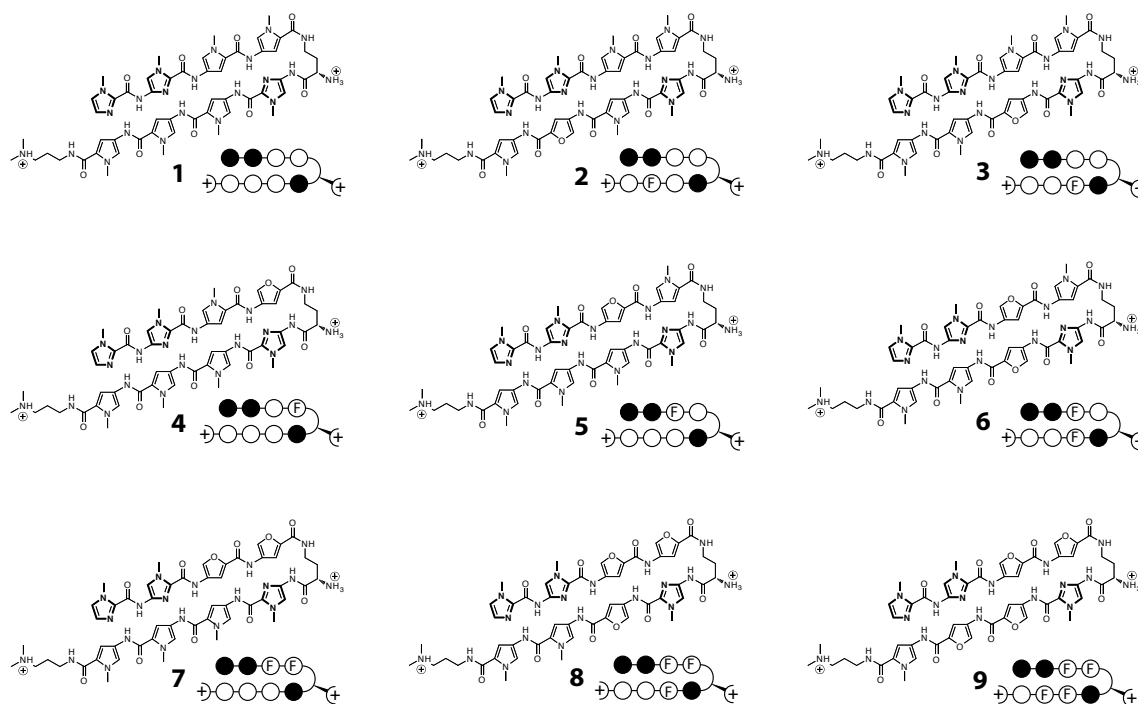


Figure A.3 Py-Im-Fn polyamide library designed to test specificity and lipophilicity

Compounds 2-9 are based on the Im-Py scaffold 1. 2-6 allow for the measurement of specificity of Fn rings in the context of polyamide pairing rules. 7-9 allow for the measurement of lipophilicity in the context of increasing furan monomer content.

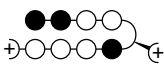
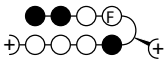
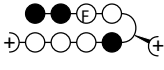
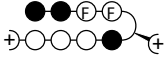
The synthesis of the Boc-protected furan monomer unit reported by Süssmuth³ is robust and we were able to report similar yields and purity in the synthesis of this monomer unit. Unfortunately, incorporation of the Fn monomer unit into Py-Im polyamides by solid-phase synthesis on oxime resin was not trivial, as the furan monomer was sensitive to the acidic conditions used for Boc removal. Excessive decomposition of the growing polymer chain was encountered following each trifluoroacetic acid deprotection. Attempts to use Lewis acid deprotection strategies (boron trifluoride diethyl etherate, trimethylsilyl chloride) were more successful, however, successive exposure of the oxime resin to the Lewis acid conditions appeared to destroy the oxime resin itself and resulted in decomposition or loss of the polyamide from the resin. Critically, polyamides containing Fn rings, once synthesized, appear to be as stable to light and heat as typical Py-Im polyamides

and remain pure by HPLC analysis even after being in solution at room temperature for several days. Because compounds **4**, **5**, and **7** require the least cycles of exposure of the furan monomer to acidic conditions, small amounts of these compounds were successfully obtained following repeated preparatory HPLC purification of the crude product cleaved from resin. These compounds along with the parent compound **1** were analyzed for binding affinity and lipophilicity.

II. Assessment of relative binding affinities and specificities of partial library

Polyamides **1**, **4**, **5**, and **7** were screened for favorable binding affinities and specificities by melting temperature analysis of duplex DNA⁴ containing the sequence 5'-TGGXCA-3' where X was cycled through the four base pair possibilities. On each of the four

Table A.1 Assessment of binding affinities and specificities of compounds **1**, **4**, **5**, and **7**

5' – C G A T G G X C A A G C – 3' 3' – G C T A C C X G T T C G – 5'					
	A–T	T–A	G–C	C–G	
DNA only	58.0 ± 0.2	58.2 ± 0.5	59.6 ± 0.1	59.8 ± 0.3	
 1	76.2 ± 1.9 [18.2]	76.2 ± 2.3 [17.9]	71.2 ± 0.2 [11.7]	72.3 ± 0.3 [12.4]	
 4	73.8 ± 0.3 [15.8]	75.2 ± 1.5 [17.0]	69.9 ± 0.1 [10.3]	70.0 ± 0.4 [10.1]	
 5	76.6 ± 1.1 [18.6]	77.1 ± 0.8 [18.9]	70.8 ± 0.6 [11.9]	72.9 ± 0.3 [13.1]	
 7	73.2 ± 1.3 [15.2]	74.6 ± 1.0 [16.3]	68.5 ± 0.5 [8.6]	69.1 ± 0.3 [9.3]	

Melting temperature studies of duplex DNA (2 nmoles/oligo) treated with polyamides (2.0-2.4 nmoles). Averages and S.D. were calculated from at least three analyses. Values in brackets represent ΔT_m versus untreated DNA duplex.

duplexes tested, the furan-containing polyamides display similar duplex stabilization to the parent compound, indicating that the binding affinity and specificity of the polyamide is likely not adversely affected by the incorporation of furan rings. This is particularly exciting considering the severe loss of DNA stabilization demonstrated by furan analogs of netropsin.³

III. Assessment of relative lipophilicity of partial library as analyzed by $\text{LogD}_{\text{octanol/water}}$ at pH 7.4.

Small molecule lipophilicity is most commonly described by its octanol/water partition coefficient ($\text{Log}^{\text{oct/water}}$). The partition coefficient is a ratio of concentrations of un-ionized compound between the two solutions. To measure the partition coefficient of ionizable solutes, the pH of the aqueous phase is adjusted such that the predominant form of the compound is un-ionized. The logarithm of the ratio of the concentrations of the un-ionized solute in the solvents is the LogP (Eq 1).

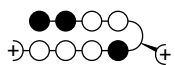
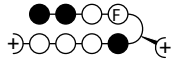
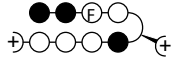
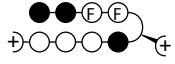
$$\log P_{\text{oct/wat}} = \log \left(\frac{[\text{solute}]_{\text{octanol}}^{\text{un-ionized}}}{[\text{solute}]_{\text{water}}^{\text{un-ionized}}} \right) \quad (1)$$

The distribution coefficient is the ratio of the sum of the concentrations of all forms of the compound (ionized plus un-ionized) in each of the two phases.. The logarithm of the ratio of the sum of concentrations of the solute's various forms in one solvent, to the sum of the concentrations of its forms in the other solvent is called LogD (Eq 2). In addition, LogD is pH dependent, hence one must specify the pH at which the LogD was measured.

$$\log D_{\text{oct/wat}} = \log \left(\frac{[\text{solute}]_{\text{octanol}}^{\text{un-ionized}} + [\text{solute}]_{\text{octanol}}^{\text{ionized}}}{[\text{solute}]_{\text{water}}^{\text{un-ionized}} + [\text{solute}]_{\text{water}}^{\text{ionized}}} \right) \quad (2)$$

Lipophilicity, as predicted by LogP and LogD, is an important predictor of oral bioavailability for test compounds. Of particular interest in the context of lead structures for animal studies is the LogD at pH = 7.4 (the physiological pH of blood serum). A variety of protocols have been reported for measuring LogP or LogD both experimentally and computationally. (5-7) The LogD and pH 7.4 of compounds **1**, **5**, and **7** was determined via a miniaturized octanol/water shake-flask assay through the commercial services of Analiza, Inc.

Table A.2 LogD at pH 7.4 of compounds **1**, **4**, **5**, and **7** at 0.1 mM concentration

		LogD	LogD*
	1	-0.95	-0.75
	2	n.d.	n.d.
	3	-0.97	-0.77
	4	-1.01	-0.80

LogD values have been corrected for DMSO background*

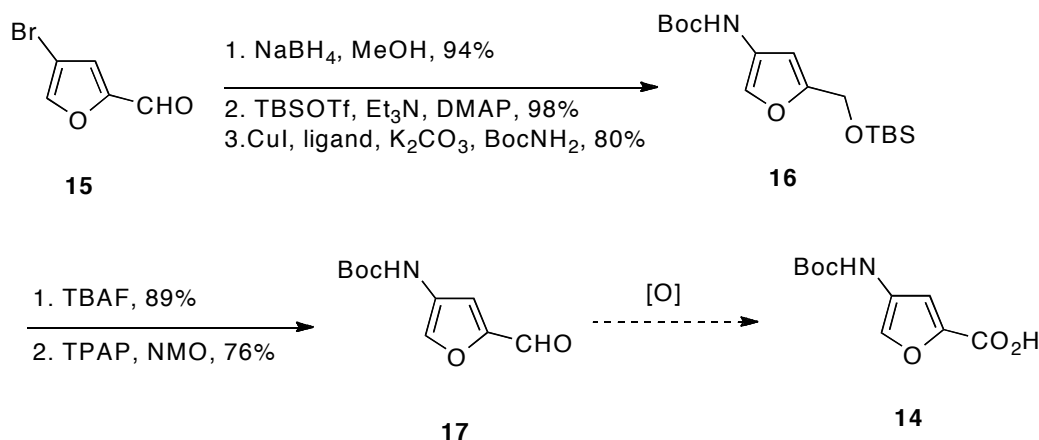
For octanol/buffer partitioning, Analiza's standard two phase system plates were used. Octanol in equilibrium with universal buffer adjusted to pH 7.4 were used to prepare partitioning plates for the assay. The initial assay was designed to measure the LogD of the compounds starting with a 10 mM stock solution, with a final working concentration of 1 mM. At this concentration, the compounds displayed significant aggregation in the octanol layer, a behavior often noted in compounds with higher molecular weights. To compensate, the compounds were diluted to 1 mM stock solution, providing a final working concentration of 0.1 mM. At this concentration, the compounds were able to be measured with the exception of compound **4**, which still displayed aggregation. Unexpectedly, the decrease in LogD over this series of increasing furan content is quite modest.

A.3 Future Directions

The binding measurements of the compounds synthesized show promise, however the modest decrease in lipophilicity observed is disappointing. It may be that compound containing a large percentage of furans will show more promise, thus the remaining compounds of the library should be synthesized. Considering the instability of the furan monomer unit to the acid conditions required by Boc peptide chemistry, a promising route is to instead utilize Fmoc-based peptide chemistry. This alternate protecting group methodology has been successfully utilized in synthesizing Py-Im polyamides and we suspect will be much milder on the furan monomer during synthesis.⁸ Once the entire library has been synthesized, lipophilicity measurements will be performed and, following this, the compounds will be tested in a protein-binding assay to determine if this alteration has the desired effect of decreasing the protein binding characteristic of Py-Im polyamides.

A.4 Synthetic Efforts

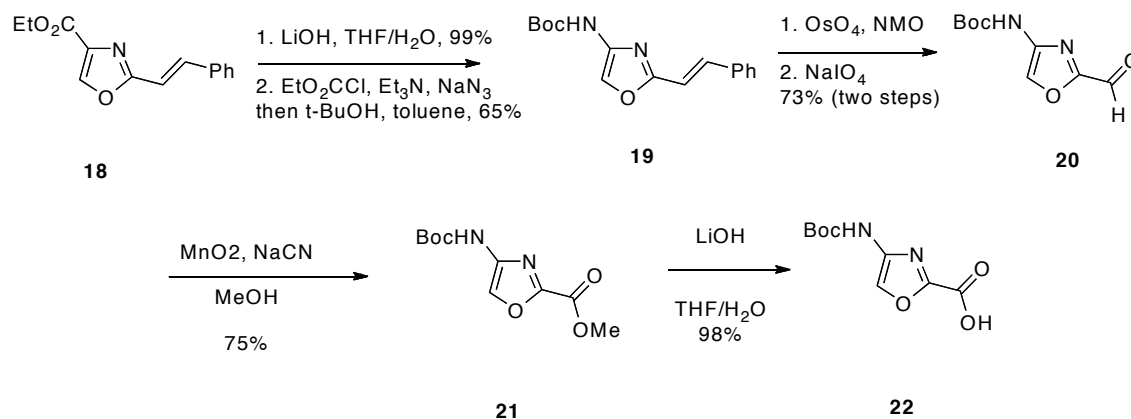
At the beginning of this project, the synthesis of furan amino acid **F** had not been reported. Accordingly, we set out to synthesize this monomer; a summary of our progress is shown in Scheme 2. The synthesis, developed by D. Gubler, began from commercially available 4-bromo-2-furaldehyde **16** and converted to carbamate **17** in three steps. The key step of this sequence, and of the entire synthesis, was a copper-catalyzed coupling of the 4-bromofuran intermediate with Boc carbamate. Unfortunately, the copper-catalyzed coupling reaction did not work on starting material **16** or its carboxylic acid derivative. Consequently, the coupling reaction was performed at the alcohol oxidation state, which, unfortunately, required two steps for reduction and protection of the alcohol followed by two steps for oxidation to give aldehyde **18**.



Scheme A.1 Attempted synthesis by D. Gubler of furan monomer Fn.

About the same time we obtained aldehyde **18**, the synthesis of 4-aminofuran-2-carboxylic acid **F** was reported by Süssmuth along with the total synthesis of the proximicins.³ This synthesis of Fn proved to be more amendable for rapid, gram scale synthesis than our route.

In addition to the furan monomer analog of Py, 4-amino-1,3-oxazole-2-carboxylic acid (Ox) has been proposed as a more hydrophilic analogue of Im. D. Gubler developed a synthetic route to the oxazole amino acid monomer which is very robust and can be done on multi-gram scale. The synthesis of oxazole monomer **22** starts from known oxazole **18**.⁹ Saponification of oxazole **18** followed by Curtius rearrangement gave Boc carbamate **19** in 65% yield. Dihydroxylation of Boc carbamate **19** using osmium tetroxide followed by cleavage of the resultant diol with sodium periodate furnished aldehyde **20** in 73% yield for the two steps. Oxidation of aldehyde **20** with manganese dioxide and sodium cyanide provided methyl ester **21**. Saponification of ester **21** smoothly furnished oxazole monomer **22** in good yield.



Scheme A.2 Synthesis of oxazole monomer Ox by D. Gubler

Unfortunately, initial attempts to couple oxazole monomer **22** into Py/Im polyamides has not been successful to date. Decomposition is seen in the coupling step under the standard coupling conditions of either PyBop/DIEA or EDCI/DMAP. Although initial forays into this area have been discouraging, more effort is needed for an accurate assessment of the feasibility of working with this monomer.

A.5 Materials and Methods

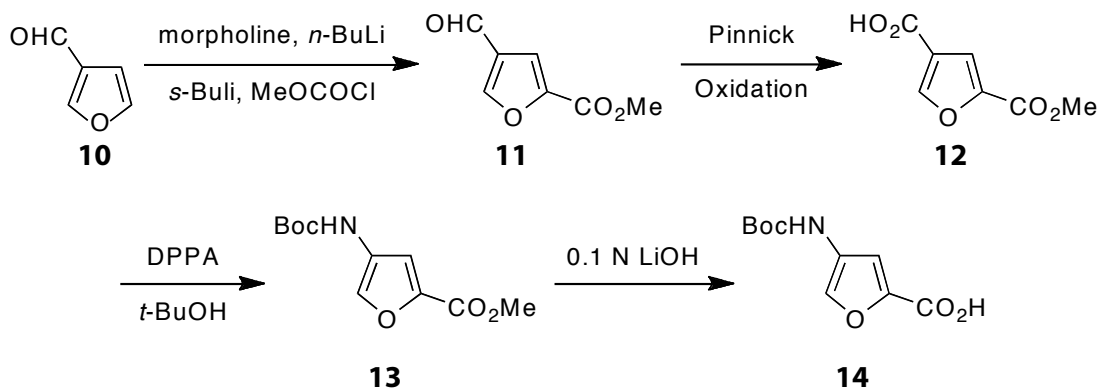
Materials

Unless otherwise stated, DNA oligonucleotides were ordered HPLC-purified from Integrated DNA Technologies. Unless otherwise stated, reagents were purchased from Sigma-Aldrich. All solvents were purchased from Aldrich or EMD Biosciences. Commercially obtained reagents were used as received. Reaction temperatures were controlled by an IKA Mag temperature modulator. Thin-layer chromatography (TLC) was performed using E. Merck silica gel 60 F254 precoated plates (0.25 mm) and visualized by UV fluorescence quenching. ICN Silica gel (particle size 0.032-0.063 mm) was used for flash column chromatography.

Methods

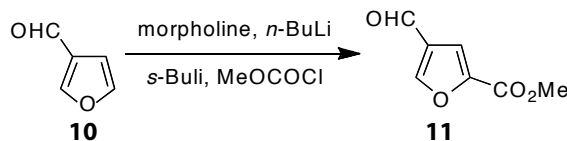
UV spectra were recorded using an Agilent 8453 UV-Vis spectrophotometer. Polyamide concentrations were measured as a solution in water at $\lambda=310$ nm using an estimated extinction coefficient of $\epsilon=69,500$ M⁻¹ cm⁻¹ for 8-ring polyamides.³⁹ LASER desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) was performed using an Applied Biosystems Voyager DE Pro spectrometer. Analytical and preparative high-pressure liquid chromatography (HPLC) were performed with a Beckman Gold system equipped with a diode array (analytical) or single-wavelength (preparative) detector. ¹H NMR spectra were recorded on a Varian Mercury 300 (at 300 MHz) or a Varian Inova 500 (at 500 MHz) and are reported relative to solvent peak. Data for ¹H NMR spectra are reported as follows: chemical shift (ppm) (multiplicity, integration, coupling constant (Hz)). ¹³C NMR spectra were recorded on a Varian Inova 500 (at 125 MHz) and are reported relative to solvent peak. Data for ¹³C NMR spectra are reported in terms of chemical shift. High-resolution mass spectra were obtained from the California Institute of Technology Mass Spectral Facility.

i. Furan monomer synthesis

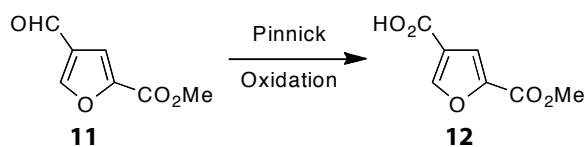


Scheme A.3 Synthesis of Boc-protected Fn monomer as reported by Süssmuth.

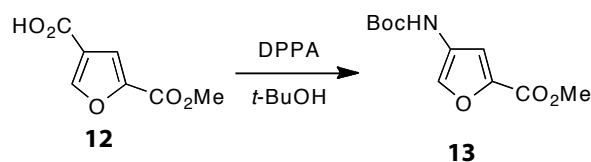
Synthesis of furan monomer was performed exactly as described by the literature procedure³:



4-Formyl-furan-2-carboxylic acid methyl ester (11). Morpholine (952 mg, 10.93 mmol, 1.05 eq) was dissolved in THF (75 mL) and cooled to -78 °C under argon. To this solution *n*-BuLi (6.830 mL, 1.6 M in hexane, 1.05 eq) was added dropwise and the resulting slightly yellow solution was stirred for further 20 min and then 3-furaldehyde (**10**, 1000 mg, 10.41 mmol, 1.0 eq) dissolved in THF (2 mL) was added dropwise. After another 20 min *s*-BuLi (8.406 mL, 1.3 M in cyclohexane, 1.05 eq) was added dropwise. The resulting suspension was stirred at -78 °C for 2.5 h. Methyl chloroformate (1082 mg, 11.45 mmol, 1.1 eq) was dissolved in THF (2 mL) and added dropwise and the reaction mixture was stirred for 45 min at this temperature an additional 40 min at room temperature. The solution was poured into ice cold 10 % HCl and the layers were separated and the aqueous phase was extracted with Et₂O (4 x 50 mL). The combined organic phases were dried over MgSO₄, filtered, and concentrated under reduced pressure to give the crude product, which was purified by column chromatography (silica gel, hexanes/EtOAc 7/1) to yield 460 mg (30%) of the desired compound. ¹H NMR δ_H (300 MHz; CDCl₃) 3.93 (s, 3H), 7.49 (d, 1H, J=0.6), 8.17 (d, 1H, J=0.8), 9.96 (s, 1H).

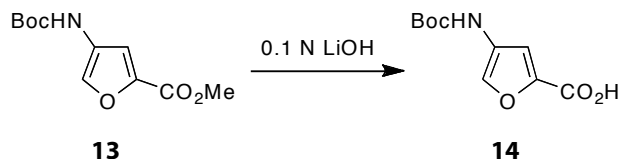


Furan-2,4-dicarboxylic acid 2-methyl ester (12). Compound **11** (860 mg, 5.58 mmol, 1.0 eq) was dissolved in *t*-BuOH (138 mL) and 2-methyl-2-butene (34 mL). 80 % NaClO₂ (5.93 g, 52.45 mmol, 9.4 eq) and NaH₂PO₄•2H₂O (6.094g, 39.06 mmol, 7.0 eq) were dissolved in H₂O (54 ml) and slowly added to a stirring solution of the aldehyde. The reaction was stirred for 50 min at room temperature. Then 2-methyl-2-butene and *t*-BuOH were removed under reduced pressure and the residue was diluted with water and extracted with hexane (3 x 50 mL). The aqueous phase was acidified to pH 2 with 10 % HCl and extracted with EtOAc (3 x 50 mL). The organic phases were combined and dried over Na₂SO₄, filtered and concentrated under reduced pressure. This slightly yellow solid was washed with CHCl₃, yielding the title compound as a colorless solid (710 mg, 75 %). ¹H NMR δH (300 MHz; DMSO-d₆) 3.82 (s, 1H), 7.40 (d, 2H, J=), 8.54 (d, 2H, J=0.83), 13.10 (brs, 1H).

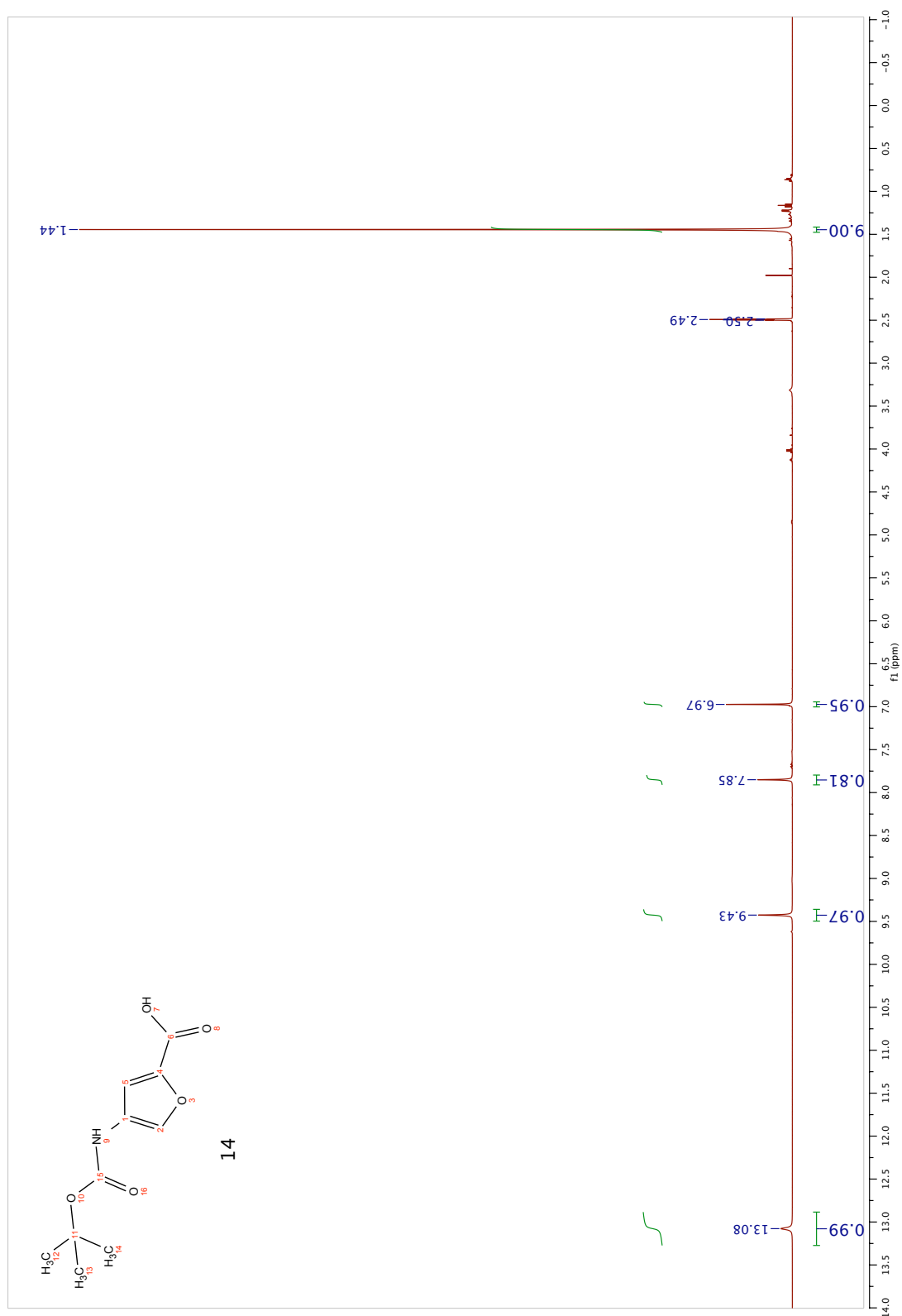


4-tert-Butoxycarbonylamino-furan-2-carboxylic acid methyl ester (13). To a solution of the acid **12** (1.3 g, 7.68 mmol, 1.0 eq) in CH₃CN (1 mL/mmol) were added sequentially triethylamine (1.069 mL, 7.68 mmol, 1.0 eq) and DPPA (1.657 mL, 7.68 mmol, 1.0 eq). This solution was stirred at RT for 3.5 h. Then *t*-BuOH (0.95 mL/mmol) was added in one portion and the resulting mixture was stirred under reflux for 24 h. The mixture was cooled to room temperature, poured into saturated NaHCO₃ (50 mL), and extracted with Et₂O (3 x 50 mL). The combined organic phases were dried over Na₂SO₄, filtered, and concentrated

under reduced pressure. The crude product was purified by flash chromatography (silica gel, 7/1, hexane/EtOAc), yielding the title compound as a slightly yellow solid (1.4 g, 78 %). ^1H NMR δH (300 MHz; CDCl_3) 1.49 (s, 9H), 3.87 (s, 3H), 6.36 (brs, 1H), 7.00 (s, 1H), 7.88 (brs, 1H).



4-tert-Butoxycarbonylamino-furan-2-carboxylic acid (14). To a solution of ester **13** (374 mg, 1.55 mmol, 1.0 eq) in THF (78 mL), LiOH (186 mg, 7.75, 5.0 eq) dissolved in H_2O (78 mL) was added dropwise. After stirring for 1.5 h at room temperature the reaction mixture was concentrated under reduced pressure and diluted with water. The solution was acidified to pH 3 by adding 10% HCl and extracted with EtOAc (3 x 40 mL). The combined organic phases were dried over Na_2SO_4 , filtered, and the solvent was removed under reduced pressure, yielding the title compound as a yellow solid (345 mg, quant. yield). ^1H NMR δH (500 MHz; DMSO-d_6) 1.44 (s, 9H), 6.97 (s, 1H), 7.85 (s, 1H), 9.43 (s, 1H), 13.08 (brs, 1H). ^{13}C NMR δC (125 MHz; DMSO-d_6) 28.46, 79.92, 111.47, 127.71, 133.60, 143.32, 153.00, 159.58. EI-HRMS m/z 227.0791 [$\text{M}^+\bullet$, $\text{C}_{10}\text{H}_{13}\text{O}_5\text{N}_1$ requires 241.0794].

**Figure A.4** ^1H NMR (500 MHz, DMSO) of compound **14**

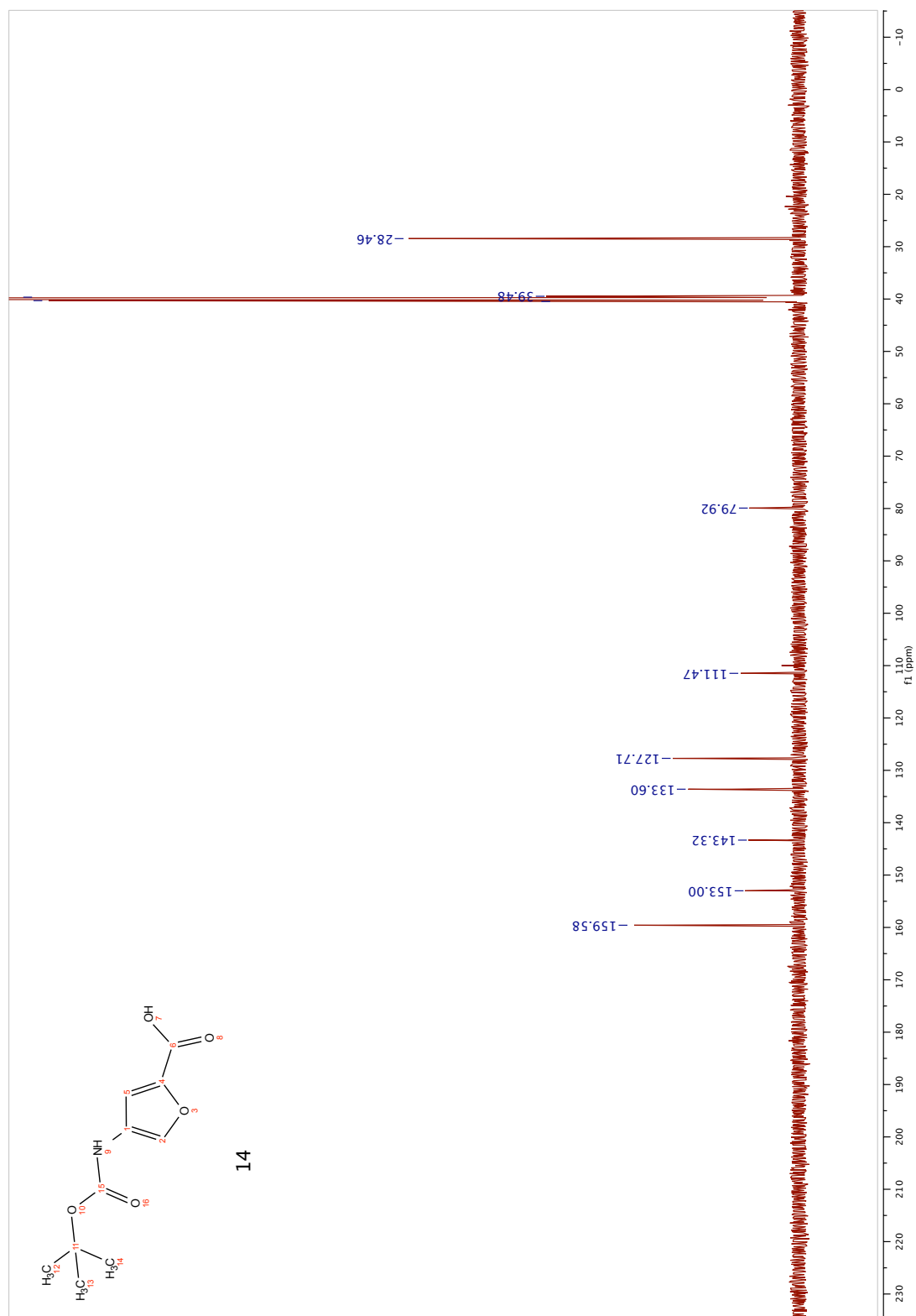


Figure A.5 ¹³C NMR (500 MHz, DMSO) of compound **14**

ii. Polyamide synthesis

Pyrole and imidazole monomer units [Boc-Py-OBt (1,2,3-Benzotriazol-1-yl 4-[(tert-Butoxycarbonyl)amino]-1-methylpyrrole-2-carboxylate) and Boc-Im-OH (4-[(tert-Butoxycarbonyl)amino]-1-methylimidazole-2-carboxylic acid) respectively] are synthesized according to established protocols¹⁰ and maintained as general group stock. Fmoc-D(Dab)-Boc-OH turn monomer unit was purchased from Peptides International.

Polyamides **1**, **4**, **5**, and **7** were synthesized by solid-phase methods on Kaiser oxime resin (Nova Biochem, Darmstadt, Germany) following established protocols,¹¹ with the exception of the Boc deprotection step. In a typical synthesis, 4 molar equivalents (relative to resin loading for the lot of oxime resin used) of monomer unit were activated with 4 equivalents of PyBOP (NovaBiochem) and 4 equivalents of diisopropylethylamine (DIEA) in a solution of dimethylformamide (DMF) (solution was approximately 0.3-0.5 M in monomer unit). In a glass peptide synthesis vessel, the resin was shaken with this solution for 2-4 hours at room temperature until the reaction was complete, as assessed by analytical HPLC of a cleaved sample of resin. Resin was deprotected by shaking in a solution of 5 equivalents boron trifluoride diethyl etherate ($\text{BF}_3 \cdot \text{Et}_2\text{O}$) in methylene chloride for 5 minutes at room temperature. After completion of synthesis the polyamide was cleaved by incubation in neat 3-dimethylamino-1-propylamine (0.3 mL) at 37 °C for 16 h. Products were purified by preparatory reverse-phase HPLC on a Beckman Gold system using either a Waters Delta-Pak 25°—100 mm, 15 μm 300 Å C18 PrepPak Cartridge reverse-phase column or a Varian Dynamax 21.4°—250 mm Microsorb 8 μm 300 Å C8 reverse-phase column in 0.1% (w/v) TFA with acetonitrile as the eluent. The appropriate fractions were lyophilized after characterization by analytical HPLC, UV-visible spectroscopy, and MALDI-TOF mass spectrometry.

1: (MALDI-TOF) $[M+H]^+$ calcd for $C_{54}H_{67}N_{22}O_9^+$ 1167.6, observed 1167.6

4: (MALDI-TOF) $[M+H]^+$ calcd for $C_{53}H_{64}N_{21}O_{10}^+$ 1154.5, observed 1154.6

5: (MALDI-TOF) $[M+H]^+$ calcd for $C_{53}H_{64}N_{21}O_{10}^+$ 1154.5, observed 1154.6

7: (MALDI-TOF) $[M+H]^+$ calcd for $C_{52}H_{61}N_{20}O_{11}^+$ 1141.5, observed 1141.7

iii. UV Absorption Spectrophotometry of DNA thermal stabilization

Melting temperature analysis was performed on a Varian Cary 100 spectrophotometer equipped with a thermo-controlled cell holder possessing a cell path length of 1 cm. An aqueous solution of 10 mM sodium cacodylate, 10 mM KCl, 10 mM $MgCl_2$, and 5 mM $CaCl_2$ at pH 7.0 was used as analysis buffer. This buffer was degassed under vacuum and DNA duplexes and polyamides were added to a final concentration of 2 μ M DNA, 3 μ M polyamide for each experiment. Prior to analysis, samples were heated to 90 °C and cooled to a starting temperature of 23 °C with a heating rate of 5 °C/min for each ramp. Denaturation profiles were recorded at $\lambda = 260$ nm from 23 to 90 °C with a heating rate of 0.5 °C/min. Each sample was subjected to denaturation in technical duplicate, and the reported numbers are the average of at least three experimental replicates. The reported melting temperatures were defined as the maximum of the first derivative of the denaturation profile.

iv. LogD Analysis

LogD analysis was performed by Analiza, Inc., information can be found at <http://www.analiza.com>. The following description regarding experimental protocol was provided by the company.

Sample Preparation:

The DMSO stock solutions were sonicated in a 40°C water bath to facilitate dissolution. Following sonication all of the samples appeared to be fully dissolved. Due to aggregation in the two phase system, the compounds were diluted to 1mM from the original 10mM DMSO stock and analyzed further. Note that for compounds **5** and **7**, some material accumulated at the interface of the octanol/water partition even at the diluted range.

Partitioning:

For octanol/buffer partitioning, Analiza's standard two phase system plates were used. Octanol in equilibrium with universal buffer (composed of 0.15 M NaCl and 0.01 M each of phosphoric, boric, and acetic acids, adjusted to pH 7.4 with NaOH) were used to prepare partitioning plates for the assay. This buffer provides uniform ionic composition across a wide pH range. DMSO stock solutions (25µL) were added to each partitioning plate to a final concentration of 10% DMSO. The plates were sealed, vortexed on our specially designed deep well plate mixer, and centrifuged to aid in phase settling. The assay was conducted on the ADW workstation using chemiluminescent nitrogen detection.

Three separate on-board performance indicating standards were assayed with the Caltech supplied compounds and all results were within the acceptable range.

Calculation of Results:

The equimolar nitrogen response of the detector is calibrated using standards which span the dynamic range of the instrument from 0.08 to 4500 $\mu\text{g/ml}$ nitrogen. Both the top and bottom phases were quantitated with respect to this calibration curve and the logarithm of the ratio of the concentration in the top phase to the concentration in the bottom phase is calculated as the partition coefficient. In addition to reporting the directly observed Log D value, the observed Log D value was adjusted to a corrected Log D* based upon our previous work correlating Log D in the presence and absence of a fixed amount of DMSO in the partitioning system.

The calculated Log D and Log D* values are corrected for any background nitrogen in the DMSO. The Log D results presented assume that the samples were free of nitrogen containing impurities and are stable under the assay conditions.

A.6 References

- (1) Chenoweth, D. M. ; Harki, D. A.; Phillips, J. W.; Dose, C.; Dervan, P. B. *J. Am. Chem. Soc.* **2009** *131*, 7182.
- (2) Fiedler, H. -P.; Bruntner, C.; Riedlinger, J.; Bull, A. T. *et al.*, *J. Antibiot.* **2008** *61*, 158.
- (3) Wolter, F. E.; Schneider, K.; Davies, B. P.; Socher, E. R. *et al.*, *Org. Lett.* **2009** *11*, 2804.
- (4) Pilch, D. S.; Poklar, N.; Gelfand, C. A.; Law, S. M. *et al.*, *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 8306.
- (5) Lombardo, F.; Shalaeva, M. Y.; Tupper, K. A.; Gao, F.; Abraham, M. H.; *J. Med. Chem.* **2000** *43*, 2922.
- (6) J. Sangster, *Octanol-water partition coefficients: fundamentals and physical chemistry* (John Wiley & Son Ltd, 1997).
- (7) Smith, D. A.; Jones, B. C.; Walker, D. K.; *Med. Res. Rev.* **1996** *16*, 243.
- (8) Wurtz, N. R.; Turner, J. M.; Baird, E. E.; Dervan, P. B. *Org. Let.* **2001** *3*, 1201.
- (9) Celatka, C. A.; Liu, P.; Panek, J. S. *Tetrahedron Letters* **1997** *38*, 5449.
- (10) Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc* **1996** *118*, 6141.
- (11) Belitsky, J. M.; Nguyen, D. H.; Wurtz, N. R.; Dervan, P. B. *Bioorg. Med. Chem.* **2002** *10*, 2767.